

# “Caption It in an Accessible Way That Is Also Enjoyable”: Characterizing User-Driven Captioning Practices on TikTok

Emma J. McDonnell  
Human Centered Design and  
Engineering, University of  
Washington  
Seattle, WA, USA  
ejm249@uw.edu

Tessa Eagle  
University of California, Santa Cruz  
Santa Cruz, California, USA  
teagle@ucsc.edu

Pitch Sinlapanuntakul  
Human Centered Design and  
Engineering, University of  
Washington  
Seattle, WA, USA  
wspitch@uw.edu

Soo Hyun Moon  
Human Centered Design and  
Engineering, University of  
Washington  
Seattle, WA, USA  
msoohyun88@gmail.com

Kathryn E. Ringland  
University of California, Santa Cruz  
Santa Cruz, CA, USA  
kringlan@ucsc.edu

Jon E. Froehlich  
Paul G. Allen School of Computer  
Science and Engineering, University  
of Washington  
Seattle, WA, USA  
jonf@cs.uw.edu

Leah Findlater  
Human Centered Design and  
Engineering, University of  
Washington  
Seattle, WA, USA  
leahkf@uw.edu

## ABSTRACT

As user-generated video dominates media landscapes, it poses an accessibility challenge. While disability advocacy groups globally have secured hard-won accessibility regulations for broadcast media, no such regulation of user-generated content exists. Yet, one major player in this shift, TikTok, has a culture of user-generated, creative captioning. We sought to understand how TikTok videos are captioned and the impact current practices have on those who need captions to access audio content. Therefore, we conducted a content analysis of 300 open-captioned TikToks and contextualized these findings by interviewing nine caption users. We found that the current state of TikTok captioning does facilitate access to the platform but that a user-generated, social video-specific standard for captioning could improve caption quality and expand access. We contribute an empirical account of the state of TikTok captioning and outline steps toward a standard for user-generated captioning.

## CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in accessibility.**

## KEYWORDS

captioning, TikTok, accessibility, user-generated video



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

CHI '24, May 11–16, 2024, Honolulu, HI, USA  
© 2024 Copyright held by the owner/author(s).  
ACM ISBN 979-8-4007-0330-0/24/05  
<https://doi.org/10.1145/3613904.3642177>

## ACM Reference Format:

Emma J. McDonnell, Tessa Eagle, Pitch Sinlapanuntakul, Soo Hyun Moon, Kathryn E. Ringland, Jon E. Froehlich, and Leah Findlater. 2024. “Caption It in an Accessible Way That Is Also Enjoyable”: Characterizing User-Driven Captioning Practices on TikTok. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*, May 11–16, 2024, Honolulu, HI, USA. ACM, New York, NY, USA, 16 pages. <https://doi.org/10.1145/3613904.3642177>

## 1 INTRODUCTION

There has been a significant shift in how people consume entertainment media; moving from traditional media (e.g., TV, movies) to online, user-generated content [95]. In 2020, for example, TikTok was the most downloaded app [73] and, in 2023, was viewed 4.4 billion minutes *per day* by US adult users alone [18]. This paradigm shift has important accessibility implications. Globally, there are well-established, legally enforced captioning standards and requirements for media that appears on television [2, 36, 97] as well as emerging legislation for professionally produced content uploaded to the internet [3]. However, the accessibility of user-generated videos remains unregulated. This leaves caption users with no legally protected access to platforms like YouTube and TikTok [62] — a notable gap.

Unlike other video sharing platforms, TikTok became popular at a time when caption use has become mainstream, including among their young user base [80]. Despite no formal requirement and significant initial obstacles, TikTok creators have developed a culture of captioning content [24, 27, 47]. Though TikTok rolled out automatic captioning and a built-in closed captioning interface in April 2021 [42], many TikTok creators have adopted a highly stylized, open captioning approach that embeds captions into their videos.<sup>1</sup> Prior HCI research on captioning design has identified

<sup>1</sup>Closed captions can be toggled on and off, using a TV remote or video player settings, whereas open captions are burned into a video and will always be on screen.

strong preferences for online captioning styles that align with television standards (e.g., [13, 28, 60, 84]), but studying TikTok offers an opportunity to explore both how creative captioning practices organically emerge on a social media platform as well as how they compare to traditional captioning practices. Despite media coverage and emerging academic interest in TikTok captioning [85], there has not yet been a comprehensive study of captioning practices on the platform. As media consumption continues to shift toward unregulated, user-generated content, studying captioning practices on TikTok provides an opportunity to understand how online videos are captioned and how those approaches serve or fail caption users.

Therefore, we set out to answer the following research questions:

- (1) How is user-generated captioning implemented on TikTok?
- (2) How do choices made in generating and placing captions impact TikTok users who need captions to meet a Deafness or disability-related access need?

To address these questions, we ran a two-phase study: a large-scale content analysis and a complementary interview study. We first collected a dataset of 300 TikToks:

- 150 targeted at a *general audience*, videos TikTok would show to a user it has very little information about (i.e., in a non-personalized feed)
- 150 related to *Deafness and disability*, videos that used one of five hashtags (#Deaf, #HardOfHearing, #Neurodiversity, #Accessibility, #Disability)

We iteratively developed a codebook and analyzed our dataset, producing an overview of the current state of TikTok captioning. To contextualize this content analysis, we interviewed nine people who rely on captions to access TikTok<sup>2</sup> about their current experiences on the platform, the impact of specific captioning choices, and preferences for the future.

We identified three major dimensions of user-generated open captions: how videos represent *audio and language* in text, how captions are *styled and placed*, and how well the *content* of captions matches a video's audio. By integrating participant perspectives with our content analysis, we found that 1) the current state of audio and language coverage in captions aligns with participant preferences—speech is nearly completely captioned while music and sound effects are rarely captioned, 2) some captions' color, size, placement, and timings varied from standard expectations, often decreasing readability, and 3) non-verbatim captioning and errors, while present in captions, were often minimally disruptive, and additional content (e.g., emoji) could provide richer paralinguistic information. Notably, we found that, despite the lack of regulation, the current state of user-generated captioning on TikTok does allow caption users to meaningfully engage with the platform. However, participants still identified significant room for growth, highlighting the need for user-generated video-specific standards, along with tools that encourage more creators to not only caption their videos, but to caption them *well*, could further extend access.

In summary, our research contributes 1) a large-scale analysis of TikTok open captioning, contextualized by its impact on caption users, and 2) steps toward future standards for user-generated captioning.

<sup>2</sup>We recruited participants based on their use of captioning, rather than a specific disability or Deaf identity

## 2 RELATED WORK

We summarize related work pertaining to video accessibility, video caption design, and research on TikTok. Our research targets anyone who needs captions to access auditory content, including d/Deaf and hard of hearing (DHH), neurodiverse, and other disabled people [29, 85]. Most prior work on captioning, however, is narrowly focused on access for DHH people, and many of the findings we recount in this section have not been broadly validated.

### 2.1 Video Accessibility

Whether or not a video is mandated to be captioned depends on where it is aired. Closed captioning on American television dates back to the 1970s and, thanks to Deaf and disabled activists, became a legal standard enforced by the Federal Communications Commission (FCC) [1, 2] via the 1990 American with Disabilities Act (ADA) and 1996 Telecommunications Act [49]. In 2010, the 21st Century Communications and Video Accessibility Act modernized these provisions to require that content aired on television with captions must remain captioned if uploaded online [97]. Legal standards vary internationally [78], changing style guidance on captioning aspects like color (e.g., [4]), and using different frameworks (e.g., W3C consortium guidelines) as the basis of law [36]. Despite some fledgling efforts to mandate captions for user-generated online video, they have not been widely implemented [3].

HCI research has found that form factor and a lack of quality captions create access barriers to online video for caption users. Viewing captions on mobile or desktop devices, as opposed to on a TV, changes users' needs, and prior work has found reason to update captioning technology and standards accordingly [21, 58, 94]. Given that not all online content is captioned, researchers explored what content ought to be prioritized for high-quality captioning [16, 84], finding that online news and educational content were the highest priorities. Berke et al. [16] noted that low-priority genres (e.g., animal videos, sports) can often be consumed and enjoyed non-auditorily. Regardless of content type, Shiver et al. found that Deaf internet users are less likely to consume user-generated video content (in this case, YouTube) when the platform is considered pervasively inaccessible [84]. More recently, Li et al. [63] explored the captioning landscape of YouTube, revealing that both creators and DHH viewers struggle to generate high-quality captions for videos and discover well-captioned videos to watch. The limited volume of work that has explored aspects of online video captioning [16, 63, 84, 85] motivates our desire to understand how TikTok is captioned.

A more recent factor shaping video accessibility is the widespread adoption of automatic captions. The rapid development of automatic speech recognition (ASR) has simplified the process of generating video captions and subsequently editing them, making captioning a far easier task [63, 65, 85]. However, automatically generated captions frequently remain unedited, which Deaf activists have highlighted as problematic [33]. Identifying when ASR-generated captions are too inaccurate poses a difficult problem, partly because long-established metrics do not adequately weight the types of errors that impact DHH viewers [50, 51]. Berke et al. [15] also caution that while some caption readers can indeed make sense of inaccurate captions, the ability to identify and make sense

of errors depends on readers' literacy levels, making inaccurate captions particularly harmful to DHH readers who face language deprivation.

## 2.2 Caption Design

On top of ensuring that videos are captioned, a significant body of work has gone into determining how to best design captions on screen. In a survey of 105 DHH caption users, Berke et al. [13] found that, given a range of options, participants preferred standard caption characteristics, including colors (black and white), timings (e.g., whole lines appearing on screen), placement (bottom of the screen), number of lines on screen (2 lines preferred), and fonts (Arial, Times New Roman, Helvetica). This suggests that DHH viewers are reluctant to move away from current, highly readable caption formats. These standard styles trace their roots back to the limited technical capacity of early American captioning technology<sup>3</sup> [78]. It is standard in North America to present captions verbatim [49].

However, accessing video content via standard captions presents several known, unresolved challenges, which many alternative captioning designs have attempted to address. Differentiating between speakers while using captioning remains a pervasive problem, and researchers have explored various solutions, such as adding animations that point to the current speaker [39, 60], moving captions next to the current speaker [43, 44, 70, 77, 98], using color and emoji [10] and designing graphical displays that use a speaker's image and name for identification [93]. Another key priority is limiting visual dispersion—that is, prioritizing designs that group relevant information within the same visual field [59]. Several studies have explored the potential benefits of dynamic caption placement in aiding viewers' comprehension of video context [21, 43, 57, 59], but it has been shown to significantly disrupt how viewers watch a video [76]. Some have found a strong interest in dynamically placing captions to identify speakers [43], while others have reported cautious interest, tempered by concerns that dynamic placement could be distracting and increase cognitive load [21]. Crabb et al. [28] found that, although captioning viewers wanted captions to be placed at the bottom of the screen by default, they strongly preferred the ability to customize caption placement. While crucial for the viewing experience, there are currently no effective tools available to guide authors in placing captions without occluding important on-screen information, a problem Amin et al. [7–9] have attempted to address by developing metrics.

Researchers have also explored various techniques, including color, animation, placement, and styling, to convey information such as volume [43, 70], emotion [61, 75, 78], and the quality of sound effects [98]. Butler terms these approaches 'aesthetic' or 'alternative' captions, contrasting them with 'integral' captions that prioritize access [22]. In a series of focus groups, she found that DHH people opposed highly aesthetic captions but concluded that creative captions that "*maintain accessible qualities*" could be useful [22]. Zdenek argues, however, that targeting more aesthetic captions to poorly captioned sections of videos (e.g., non-speech sounds) could leverage the expressive capacity of approaches like kinetic typography while preserving readability [98]. Research has

also shown that more humble interventions, such as using punctuation to indicate pauses in automatic captions [40], can positively impact caption readability. Though there are documented practices of translating internet community language into subtitling, it has not been reviewed in the context of accessibility [83]. Several researchers have attempted to address the lack of captions for non-speech sounds by creating authoring tools to graphically represent sounds [5] and generating a set of representations for domain-specific sounds [25], finding it is particularly important when sounds arise off-screen [46]

## 2.3 TikTok and Research

The social internet has shifted in recent years toward short-form video content. TikTok, developed in 2016, is an algorithmically-driven social media platform primarily focused on video sharing. Since 2023, the platform has reached over 1 billion active users worldwide<sup>4</sup> and was the most downloaded app in 2020 [73]. Given this rise in popularity, other platforms introduced similar features for publishing brief video content (e.g., Instagram Reels – released in August 2020<sup>5</sup> and YouTube Shorts – released in September 2020<sup>6</sup>). Content on TikTok is primarily short-form video-based, with videos initially limited to 60 seconds in length, and more recently extended to ten minutes [91]. TikTok is not only notable in its bias toward short content, which may be easier to caption, but recent work highlights ways that TikTok's platform incentivizes specific kinds of content (e.g., a strong bias toward repetitive trends, a desire to optimize content for the algorithm) [17, 101], which can encourage a culture of open captioning. While initially adopted by younger populations for dance-related challenges, TikTok's user base has since diversified, and the app is now primarily used by 19 to 29-year-olds<sup>7</sup>.

TikTok has recently gained considerable attention in HCI research. Much of this work has centered on sensemaking around the proprietary and elusive TikTok algorithm [31, 52, 54, 71, 86] and specific sub-communities that vary widely from grieving individuals [37], to those with experiences of psychiatric hospitalization [82] and eating disorders [41], to users discussing acne and skincare [38, 100]. Marginalized groups have also found community and belonging on TikTok. For example, research has highlighted LGBTQ+ communities [31, 87], neurodivergent-related content [6, 35], inclusive technology for disabled individuals [34], and discussions of shadowbanning in queer, trans, and disabled TikTok communities [79]. Research on the credibility of information disseminated on TikTok has also proliferated, especially with regards to the COVID-19 pandemic [11, 64, 88]. Notably, these studies often analyze the 100 most liked or viewed TikToks within certain topics or hashtags [56, 81, 96, 100].

However, little work has focused on the accessibility of the platform. TikTok did not introduce automatic captioning until April 2021 [42]. Given the primarily video-and-audio-based nature of the platform, captions are an integral part of participation for d/Deaf,

<sup>3</sup>European standards differ from American standards, often using more colors and fonts, due to a different initial technical implementation [78]

<sup>4</sup><https://www.demandsage.com/tiktok-user-statistics/>

<sup>5</sup><https://about.instagram.com/blog/announcements/introducing-instagram-reels-announcement>

<sup>6</sup><https://blog.youtube/news-and-events/one-year-of-youtube-shorts-what-weve-learned-so-far/>

<sup>7</sup><https://www.demandsage.com/tiktok-user-statistics>

hard of hearing, neurodiverse, and disabled communities. Simpson et al. found that much of TikTok's accessibility has stemmed from grassroots community efforts, largely by disabled communities who have developed workarounds to address app inaccessibility [85].

### 3 METHODS

We employed a two part mixed-methods study. We first collected and analyzed a dataset of TikToks to characterize how user-generated TikTok videos are captioned. Then, we performed complementary interviews with TikTok users who need captions to access the platform to identify the impact of these current captioning approaches.

#### 3.1 Content Analysis

We began by collecting a dataset of TikTok videos and developing an initial codebook. We describe our method for collecting and analyzing videos, which led to the overview of TikTok captioning approaches we present in Section 4.1.

**3.1.1 Data Collection and Analysis.** We created a dataset comprised of (1) TikToks likely to be shown to a general audience and (2) TikToks related to Deafness and disability. We took this two-pronged approach<sup>8</sup> to understand how content is captioned both when it reaches a broad audience and when it is made by communities invested in access. We chose to collect both general audience and Deafness and disability-related videos to analyze a breadth and variety of captioning practices on TikTok. Further, our analysis was targeted at understanding current practices on the platform, independent of captioning users' viewing patterns. For this research, we define captioning as a textual representation of audio or language, including spoken language, signed languages, and other sounds.

As background, captions can be either open or closed [32]. Open captions are burned into video content, whereas closed captions can be toggled on and off. See Figure 1 for an illustration of the difference. While TikTok supports both open and closed captioning we observed inconsistency in the availability of closed captions during preliminary analysis, with variation over time (e.g., videos appeared closed captioned one day and not another) and across devices and browsers (e.g., at the same time on the same device, videos appeared with closed captions in a mobile browser and without closed captions in the TikTok app). Due to this inconsistency, we scoped content analysis to open captions.

Because TikTok's terms of service prohibit "*us[ing] automated scripts to collect information from or otherwise interact with the Services*" [92], we collected data manually. To collect data, researchers created new accounts and liked or favorited<sup>9</sup> videos that met each data collection phase's inclusion criteria. After liking and favoriting the quota of videos for each data collection round, researchers requested their account's data from TikTok. This resulted in a JSON file containing the links to all videos that a user had liked and favorited. We parsed and combined these files, using the resultant list of video links to form our dataset.

<sup>8</sup>Other accessibility research has also used multiple data sources to gain a more complete picture of the area of study (e.g., [67])

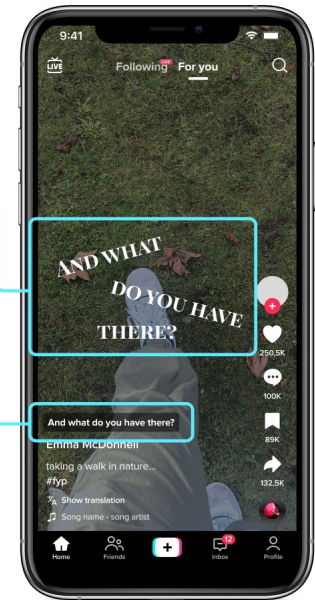
<sup>9</sup>These functions serve equivalent purposes, but we observed occasional malfunctions with the "Like" feature and used "Favorite" as a backup

#### OPEN CAPTIONS

Open captions are "burned" into the video itself and added by the creators with their own textual style and placement

#### CLOSED CAPTIONS

Closed captions always appear at the bottom of TikTok videos with white on black text and can be auto-generated



**Figure 1: Simulated screenshot of a TikTok illustrating the difference between open captions (top text of the video) and closed captions (bottom of video). Closed captions appear at the bottom of a TikTok video as white text on a translucent black background and can be toggled on and off. Open captions can be any color, size, font, and in any location on the screen and are permanently part of a video.**

**3.1.2 General Audience Data Collection.** To characterize broader trends in TikTok captioning, we sought videos that were likely to be shown to a broad audience. However, TikTok's emphasis on personalized, automatically generated video feeds (a 'For You Page' or FYP), means there is not a core set of videos all TikTok users see. Therefore, we targeted videos that TikTok serves a user it has very little information about, as a proxy for general audience data. To collect this data, four members of the research team generated new TikTok accounts, and, over five days in early February 2023, each researcher liked or favorited 100 captioned videos a day. We scrolled through the research account's FYP, liking a video if it was captioned and scrolling as soon as we determined it was uncaptioned. We excluded ads, live videos, and sponsored posts from consideration but had not yet discovered inconsistencies with closed captions, so we collected both open and closed captioned videos.

We initially collected 2000 general audience videos. Among these, 1654 were unique URLs, signifying unique videos. At the time of submission, 65.3% (1050) of the 1654 unique videos featured open captions, 28.1% (464) had no open captions, and 8.5% (140) had been taken down since initial collection. Having intentionally over-sampled, researchers then randomly selected 150 videos from the set of 1050 open-captioned videos for coding and analysis, informed by sample sizes in prior work (see 2.3).

**Table 1: Frequency of Deafness and disability data collection hashtags in the final dataset**

Hashtag	Count
#Disability	49
#Deaf	44
#Accessibility	30
#Neurodiversity	29
#HardOfHearing	22

**3.1.3 Deafness and Disability Data Collection.** We also collected data from communities we hypothesized to be at the cutting edge of video accessibility—Deafness and disability-related content creators. We identified Deafness and disability-related videos via the following five hashtags: #Deaf, #HardOfHearing, #Disability, #Accessibility, and #Neurodiversity. We selected these hashtags to balance gathering videos with a broad focus (e.g., #Disability, #Accessibility) with videos targeting communities likely to use captions to access videos (e.g., #Deaf, #HardOfHearing, #Neurodiversity) [63, 85]. We sought this balance to ensure we had representation from communities that value captioning while not excluding groups we did not consider in advance. To collect this data, we generated five new TikTok accounts, which were used to collect data by the same four researchers who collected the general audience data (the lead author collected data on two accounts during this data collection cycle). Each account was assigned a different hashtag to collect data from daily, assigned over a five-day period. We used a Latin Square design to ensure that 1) each research account was used to collect data exactly once from each hashtag and 2) we collected data from every hashtag for each of the five days of data collection.

Over five days in April 2023, researchers searched their designated hashtag on the TikTok 'hashtags' results tab and scrolled until they had liked or favorited 100 open-captioned videos daily. These hashtags often had a high concentration of videos from a small set of creators and sometimes contained irrelevant or offensive content. Therefore, while collecting data researchers strategically avoided liking videos from the same creator to diversify our dataset and excluded content they deemed irrelevant. Researchers were instructed to skip a video if it was 'clearly non-topical', 'ableist mockery', or an ad, and borderline videos were included to be discussed later. The data collection process was designed to collect 2500 videos, with 500 per hashtag. However, one researcher's device did not consistently register 'liked' videos and was only able to record 237 out of 500 'liked' videos. We, therefore, collected 2,237 videos, 1,208 of which were duplicates, resulting in a final dataset of 1,029 videos. To match our general audience data, we coded and analyzed a random set of 150 of these videos. The distribution of hashtags in our final dataset is shown in Table 1.

**3.1.4 Video Content Analysis.** We iteratively developed a coding scheme to analyze how videos are captioned. Over three cycles, four researchers drafted a set of codes, applied them to 25–30 videos, and discussed gaps, redundancies, and disagreements before settling on a coding scheme. After the final round of coding, researchers achieved an average Krippendorff's Alpha inter-rater reliability

score of .71 on all quantitatively analyzed codes.<sup>10</sup> The final coding scheme tracked three key components of captioning: audio and text coverage, style and placement, and caption content.

We then applied our coding scheme to 150 general audience and 150 Deafness and disability TikToks. To diversify our examination of captioning practices, we analyzed only one video per creator. The same four researchers who collected data and generated the coding scheme coded the videos, with two researchers coding each video over the course of two rounds. In the first coding round, each researcher coded 75 open captioned videos (one half General Audience, one half Deafness and Disability Related). Each coder's set of videos was then randomly sorted into thirds and distributed to other members of the coding team. During the second round, each researcher again coded 75 videos. Upon completing both rounds of coding, each pair of researchers discussed and resolved the differences between their coding of the 50 videos they both analyzed. This process produced a single, authoritative coding for each of the 300 videos we analyzed.

We then performed a mixed-methods analysis of our coded data. For quantitative data, we calculated summary statistics and for qualitative data, we open coded responses<sup>11</sup>. To conduct this analysis, we split our coding scheme into thirds and had two researchers analyze each third, with the lead researcher taking part in all analyses.

We also conducted a word error rate (WER) analysis on the videos we identified as containing at least one error. Though we identified errors in 59 videos, only 55 were still posted on TikTok at the time of calculation. For those videos, we transcribed the open captions directly and then manually generated a verbatim transcript of the video. We used the Amberscript implementation of the NIST Scoring Toolkit<sup>12</sup>, to calculate the WER for each video and computed the overall average WER. We also calculated the WER for the three videos shown during the interviews.

## 3.2 Interview Study

To complement our TikTok video content analysis, we performed semi-structured interviews with TikTok users who need captions to access the platform, seeking to understand the impact of common TikTok captioning approaches. Following Mack and McDonnell et al. [68], we defined eligibility by captioning use, rather than specific disability, recruiting participants who use captions "*due to Deafness, disability, neurodiversity, or related condition.*" We relied on established connections within Deaf and disability communities to recruit participants, reaching out to relevant mailing lists and using snowball sampling.

The semi-structured interviews, conducted over video conference, lasted one hour and had three parts (See Supplementary Materials for the interview protocol). First, we asked participants to reflect on their current experiences with TikTok and how (in)accessible videos are to them. We then selected three videos, all with more than 500,000 likes<sup>13</sup>, from our dataset, which exemplified key aspects of

<sup>10</sup>A full codebook, granular IRR data, and codes contained in this group can be found in the supplementary materials

<sup>11</sup>Supplementary materials include specifics on how each characteristic was analyzed

<sup>12</sup><https://www.amberscript.com/en/wer-tool/>

<sup>13</sup>We intentionally selected videos with a large viewership to mitigate privacy concerns



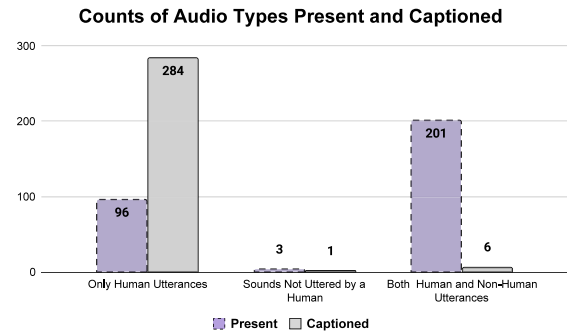
captioning identified in our content analysis. The first video<sup>14</sup> (WER = 6.1%) captioned speech but not background music and used varied caption color, placement, and size. The second video<sup>15</sup> captioned one of the two speakers but did not caption the dog – a salient audio source – or the yelling by the two speakers. Participants were shown the original video (WER = 37.5%) as well as an edited version made by the first author, which captioned all audio (WER = 0.0%). The final video's<sup>16</sup> captions (WER = 0.0%) were formatted standardly but represented a voice-over track not connected to on-screen actions. After participants viewed the videos, we probed for their reaction to certain aspects of the captioning. Finally, the study session concluded with a discussion of what participants would like to see in the future and a comparison of current TikTok norms to other video content. Participants were compensated \$40, automatic captions were always enabled, and we arranged ASL interpretation and CART transcription upon request.

We had nine participants in this study. Their average age was 39.1 years old (range 19-73), and five identified as Deaf, three as deaf<sup>17</sup>, two as hard of hearing, one as neurodiverse, and two as having some other disability (some participants held multiple identities). With regard to gender, seven participants identified as women and two identified as men. Participants self-reported their race: 66% were white, 11% Black, 22% Asian or Pacific Islander and 11% Native American. We required that participants frequently use TikTok, and 44% reported using the platform multiple times daily, 22% reported daily use, 22% reported using TikTok 3-5 times a week, and 11% reported weekly use. We required participants have experience reading captions in English and five also reported communicating using ASL.

We used a mix of top-down coding and reflexive thematic analysis [19, 20] to analyze interview data. Upon completing interviews, researchers reviewed transcripts, flagging data aligned with content analysis findings and taking notes to form a codebook.<sup>18</sup> Data that aligned with content analysis findings was open coded and integrated accordingly. Researchers coded the remaining data in two stages - one researcher completed the initial coding pass and a second reviewed their work. Across this process, the lead researcher reviewed all transcripts. Coded data was then developed into themes using an inductive, semantic, and critical realist approach. Thematic analysis emphasizes the role of authors' positionality. We are a mixed-ability research team, with some members identifying as DHH, neurodivergent, and/or disabled. The lead researcher is a hearing person with conversational ASL skills. Authors identify as white and Asian.

## 4 FINDINGS

We present a content analysis of TikTok captions, highlighting key considerations that go into captioning and integrating participant perspectives on the impact of different captioning choices. We then identify broader themes around the state of TikTok accessibility for captioning users.



**Figure 2: Bar chart displaying the number of videos where each audio type was present and videos that captioned each respective audio type. Videos with only human utterances accounted for the majority of captioned videos (94.7%, 284/300)**

### 4.1 The Current State of TikTok Captioning

To characterize TikTok captioning, we examine how audio and language are represented in caption text, style and design choices, and caption content.<sup>19</sup> As relevant, we contextualize these video analysis findings with interview participants' perspectives and preferences. Table 2 provides an overview of the facets of captioning we analyze in this section.

**4.1.1 Audio, Language, and Text.** We required videos in our dataset have captions, but creators did not caption all audio equally. Understanding what audio is present in videos and how comprehensively it is captioned reveals what creators prioritize when making a video accessible. Therefore, we focus on *how much* audio is captioned in TikTok videos and how that aligns with participant preferences. We provide an overview of audio types then discuss how human speech, music, non-speech sound, and signed languages were captioned.

We categorized video soundscapes as containing 1) human utterances only (e.g., speech, singing, laughter), 2) sounds not uttered by a human (e.g., dog barking, instrumental music, clapping, appliance beeping), or 3) a mix of both. Figure 2 shows how often each audio type occurred in our dataset and how frequently each type was captioned. Most, but not all, videos (72.7%, 218/300) consistently captioned each audio type. Videos were considered to be captioned consistently in two situations – when *all* of one audio type was captioned in videos that audio type, or if an audio type was present and *never* captioned. For example, if a video contained spoken and sung human utterances and did not caption singing, we considered its captioning inconsistent.

**Human Utterances.** Human utterances were part of nearly every video in our datasets (99%, 297/300), and were largely captioned (96.7%, 290/300). Most commonly, these captioned human utterances were speech – 85.7% of videos (257/300) contained people

<sup>14</sup><https://www.tiktok.com/@austinandlexi/video/7188243037972106539>

<sup>15</sup>[https://www.tiktok.com/@bananna\\_k/video/7198305835943185710](https://www.tiktok.com/@bananna_k/video/7198305835943185710)

<sup>16</sup><https://www.tiktok.com/@ripleysaquariums/video/7167494942204497157>

<sup>17</sup>Capital "D" Deaf often signals identity with Deaf community, whereas lower-case "d" deaf more frequently refers to the audiological experience of deafness [72]

<sup>18</sup>See Supplementary Materials for the codebook

<sup>19</sup>See the Content Analysis Codebook in Supplementary Materials

**Table 2: An outline of the key facets of user-generated open captions that we discuss in Section 4.1**

Category	Facets to Consider
Audio, Language, and Text	<ul style="list-style-type: none"> <li>- Types of sounds in videos</li> <li>- Amount of audio covered by captions</li> <li>- Approaches to captioning signed languages</li> </ul>
Style and Placement	<ul style="list-style-type: none"> <li>- Caption timing and animation</li> <li>- Caption placement and alignment</li> <li>- Use of color</li> <li>- Text formatting</li> </ul>
Content	<ul style="list-style-type: none"> <li>- Deliberate non-verbatim transcription</li> <li>- Additional content</li> <li>- Captioning accuracy</li> </ul>

talking, and only five did not caption all speakers. This prioritization matched participant preferences: they unequivocally agreed that uncaptioned speech was inaccessible. As P1 put it, *“no captions, and I just scroll past”*.

**Music.** Though many videos in our dataset contained lyrical or instrumental music, it was often uncaptioned—interestingly, a decision participants supported. Though we could not reliably quantify the presence of music in videos, we observed that it was rare for the presence of music (instrumental or lyrical) to be indicated in captions. When captioned at all, lyrics were often selectively captioned rather than fully transcribed. Largely, participants did not find captioning music to be necessary on TikTok. Many participants agreed that it was *“honestly easier not to know”* (P9) about most music because, on a small screen, *“it just gives more things that I need to read and then it gets frustrating”* (P6). Some participants linked this to their Deaf identity: *“music does nothing for me, I don’t understand it—I am profoundly Deaf and have always been Deaf”* (P4). If music is captioned, participants preferred that creators use a music note emoji (P2) or briefly note the tone indicated by the music (e.g., [upbeat music] P8) rather than transcribing lyrics.

**Signed Languages.** We observed several videos that captioned sign language (13.3%, 40/300), all from our Deafness and disability data collection, though our interview participants reported infrequently coming across videos that captioned signed languages. Signed videos took a wide range of approaches to audio: 22.5% (9/40) captioned no audio and only captioned signing. However, often people signed and spoke content simultaneously<sup>20</sup>, interpreted music and TikTok sounds, or used text-to-speech to voice an English interpretation of their signing. However, because our data collection processes sought captioned videos, they likely do not reflect all signed content on TikTok. P6 and P8 both reflected on encountering uncaptioned signed videos as DHH people who know some ASL but do not primarily sign, noting it is *“kind of weird for me, because I’m like, you want me to understand you, but you’re going to make me work for it”* (P6). P9 pointed out that captioning signed videos poses a challenge, as there is not a *“standard way to have captions for our language.”*<sup>21</sup>

**Non-Human Utterances.** Only 2.3% of videos (7/300) in our dataset captioned non-human utterances, a stark contrast to how

spoken and signed languages were captioned but somewhat aligned with participant preferences. Interview participants were mostly interested in captioning ‘important’ sounds and sounds that were not obviously visually indicated. Participants stressed the importance of considering the purpose and impact of sound in a video: if *“someone’s just making a like, kind of annoying, stupid noise, I don’t really need context for the noise they’re making”* but captioning relevant sounds *“added flavor to the video”* (P6). Additionally, P9 noted that the TikTok format made environmental audio less relevant than in other media: *“if you miss sound on a [TikTok] video, you can still enjoy it, but for movies you are left wondering.”*

**4.1.2 Style and Placement.** TikTok captions are notable in their use of a wide range of approaches to style and placement. As P3 put it: *“I think that the captions on TikTok are way way way more creative and people seem to be having more fun with captioning compared to [traditional video platforms].”* We sought to understand how videos in our dataset approached caption timing and animation, placement, color, and formatting. Overall, while there was nonstandard style and placement throughout our dataset, participants preferred captions that prioritized practical access over novel designs.

**Timing and animation.** When choosing how to time and animate captions, the majority of videos aligned with participants’ preferences for captions to be *“static, right there, simple, clean”* (P1). Most videos (83.3%, 250/300) timed their captions similarly to movies and TV: a few lines appear on screen at a time and refresh once all content is spoken. Other timings included captioning speech one or a few words at a time (5.0%, 15/300 videos) and emulating live captions (3.3%, 10/300), with words appearing as they are spoken, building into captioned lines (see Figure 4). Most participants stressed the need for captions to not disappear *“so quickly that I don’t have time to read it”* (P2). P9 provided a differing perspective, noting that while rapid-fire captions are *“not 100% accessible”*, she *“really like[s] it, cause it shows me the way [a TikTok creator is] talking.”* Caption rate has long been considered an obstacle to caption readability [48], and user-driven choices around caption timing add another dimension to this discussion. 10.0% of videos (30/300) animated captions in some way, occurring more frequently in general audience videos (15.3%, 23/150) than in Deafness and disability videos (4.7%, 7/150). Common animation styles include fading, bouncing, and erratic motion (e.g., strobing, shaking) (see

<sup>20</sup>Known as simultaneous communication or sim-comming

<sup>21</sup>English language captions can never directly represent signing and only provide a written interpretation [12]

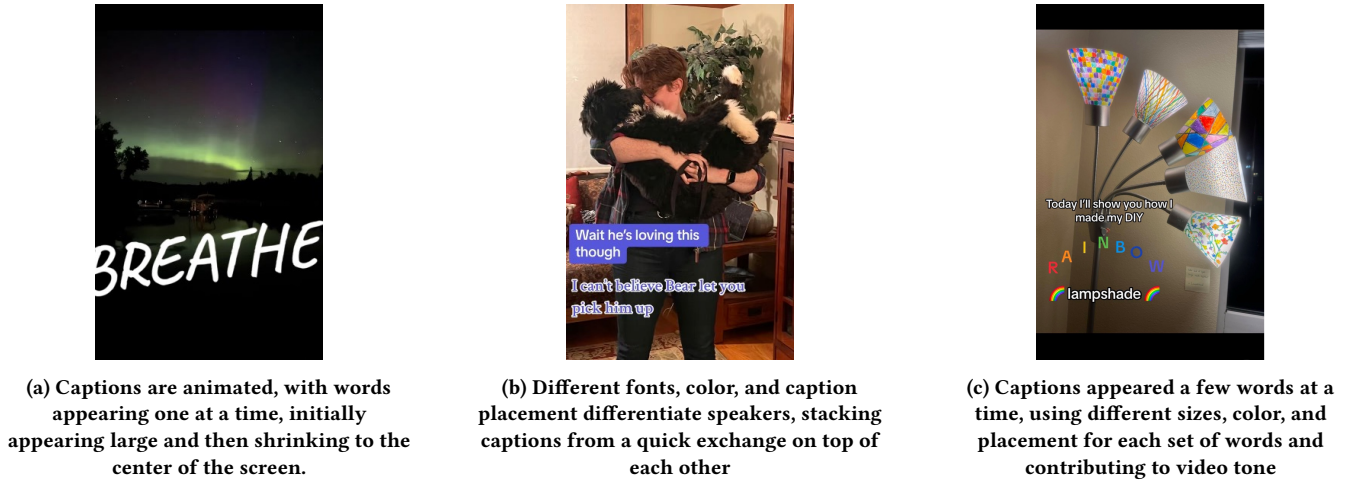


Figure 3: Simulated stills representing aspects of caption style and placement we observed throughout our dataset

Figure 3). Participants noted that this amount of motion on screen “can be really jarring” (P3).

**Placement.** Despite interview participants’ strong preference for captions that stay in one location, over a third of videos (34.3%, 103/300) moved captions around the screen over the course of the video. Variable caption placement was more prevalent in Deafness and disability videos (43.3%, 65/150) than in general audience videos (25.3%, 38/150). Commonly, caption placement was used to differentiate information—for example, separating types of audio (e.g., TTS, laughter, human speech) or contributions from multiple people (see Figure 3). Caption movement could also serve as a meta-structure to organize the video’s content (e.g., separating questions and answers, moving from topic introduction to content). However, many videos included seemingly random placement or placement motivated solely by a high-contrast background. Captions were placed in all regions of the screen, with a slight bias toward the top of the video, a departure from established practice [21].

Participants consistently reported problems with poorly placed captions. Often, TikTok’s dense UI elements on the bottom and right sides of the screen overlap with captions and make it so “I can’t see those captions” (P9). In contrast to many captioning standards, P7 suggested that creators should default to placing captions along the top of the video as “more things are happening on that floor 80% of the screen instead of like the top 20%.” Additionally, participants did not like when the “captions felt far from the action” as their “eyes were doing double work, popping up and down” (P4), a common captioning consideration known as visual dispersion [59]. However, the value of placing captions near relevant visuals comes into conflict with the desire to not move captions around the screen. If captions move throughout a video, P2 noted “I had to look all around to figure it out ... If it was all in one place each time, then I know where to look for placement.”

**Color.** Color choice had a strong impact on caption comprehension, including both the text outline and fill colors. Over 87% (262/300) of videos used black-and-white captions (see Table 3), which were preferred by our interview participants. However, 29.3% (88/300) used other color combinations, most commonly white text

Table 3: Frequency of open caption color scheme in our dataset

Caption Style	Example	Frequency
White with Black Outline	Example	54.7% (164/300)
White with No Outline	Example	15.3% (46/300)
Black with White Outline	Example	3.7% (11/300)
Black with No Outline	Example	2.7% (8/300)
White on Gray/Black Background	Example	9.3% (28/300)
Black on White Background	Example	10.3% (31/300)
White with Colorful Background	Example	15.7% (47/300)
Colorful with White Outline	Example	7.7% (23/300)

on a colorful background (15.7%, 47/300 videos) or colorful text outlined in white (7.7%, 23/300). Videos used multiple captioning colors 22.3% (68/300) of the time, which most frequently served to differentiate speakers and sounds or to emphasize specific phrases within the video (see Figure 3). Participants’ priority for color schemes was that they produce “simple captions that I can read” (P5), allowing for more colorful captions to differentiate speakers or sound sources only if readability was the guiding principle. Suggestions included using bright color in the caption background and keeping the text black (P6) or, as P3 suggested, using different combinations of black and white: “maybe black text with white background for the first speaker and black background and white text for the second speaker, so that way it feels more consistent”.

**Formatting.** Videos also leveraged formatting elements such as typeface, size, and capitalization to style their captions, but most



deviations from a perceived norm were not well-received by participants. A vast majority (94.7%, 284/300) of videos used the same typeface throughout their video, but variation was used to differentiate video titles from captions, to emphasize the final lines of videos, and to indicate a speaker change (see Figure 3). We also observed that videos frequently departed from standard rules of capitalization, using all-caps to emphasize certain words, captioning some videos entirely in all-caps, and sometimes not capitalizing any words in captions. Participants did not find this to be helpful variation: *“I just don’t want them changing the style and the font and the letters—that is really hard”* (P5). Though we were unable to consistently quantify trends in font size, we observed a high degree of variation. Size changes could be meaningful, differentiating an important phrase from the rest of captions, or somewhat randomly scaled with regard to the amount of text on screen. Participants suggested that having the font *“large enough to be able to read”* (P2) is critical to readability and noted that if they come across *“captions that are like, tiny ... I can’t read that ... I’d swipe through [and skip the video]”* (P8).

**4.1.3 Considerations Around Caption Content.** While the above two subsections focused on how audio was translated into captions and how captions were styled and placed, here we consider the content of captions themselves. Going beyond the presence of captions and considering their content, we analyzed instances when language was not captioned verbatim, when additional content was added, and when the captions had errors (see Figure 4).

**Deliberate Non-Verbatim Captioning.** Though the Deaf community has long advocated for verbatim captions [49]—as opposed to summarized or censored captions—our interview participants had nuanced, context-dependent perspectives on non-verbatim, user-generated captions. In our video analysis, we found that 18.3% (55/300) videos deliberately used non-verbatim captions, most often to caption a curse word or other vulgarity (e.g., “shit” captioned as “sh’t”) though sometimes to replace content that was not obviously a censorship target (e.g., “autism” captioned as “the ‘tism”). Popular strategies for altering words included using asterisks or other punctuation in place of vowels, removing letters from words, replacing the spoken word with an alternative (e.g., “fucking” captioned as “friggin”), and using acronyms, abbreviations, and emoji in the place of fully voiced words. Notably, these instances of non-verbatim transcription were limited in scope, often impacting single words in videos. Participants largely echoed P6’s reaction that they *“haven’t come across censored captions too much”*, but that *“it’s kind of just like something you have to deal with.”* Breaking with a long tradition of strong opposition to censored captions [99], many participants shared P9’s sentiment that, while *“I don’t like it when they [censor content] ... I understand the creator’s reasoning.”* Participants still disliked the ways that censorship feels like *“you’re treating me as if I’m less than or as if I’m fragile because I can’t hear”* (P6). However, they considered that audiences contain *“young people too”* (P8) and that creators may need to protect their content on a platform prone to censoring videos [79]. In fact, P3’s initial reaction to much of the non-verbatim content she saw on TikTok was that it *“feels like I’m getting older”* as she noticed patterns of captioning that *“kind of became a language and a culture to get around the censors.”* Overall,

while non-verbatim captions provide lower-quality access, participants took a nuanced view, understanding them as part of platform culture in the face of censorship and shadowbanning.

**Adding Content to Captions.** In our dataset, 23.7% (71/300) of videos added content to captions beyond direct transcription, often communicating paralinguistic aspects of speech (e.g., tone). This occurred more frequently in Deafness and disability videos (34.7%, 52/150) than in general audience videos (12.7%, 19/150). Most commonly, videos included emojis, frequently used to indicate the tone of the spoken content (e.g., 😊, 😭, 😏) or to match the topic of the video (e.g., 🌊, 🐳, 🎵 for a video using the song “Under the Sea”). Participants largely liked emoji additions but emphasized that while sparing emojis can *“help me understand mood and the perspective,”* excessive use is *“a little bit cringey”* (P6). P4 likened emojis that matched the tone of a caption to non-manual markers, a key component of ASL grammar that often serve as a tone modifier. While many videos entirely omitted punctuation, when used, punctuation helped to differentiate types of content (e.g., indicating that *“whispers”* was a tonal description, not a captioned word), to convey volume or emphasis (e.g., using !!! and !?), and to convey the pace of speech in captions (e.g., a caption that reads *“It’s just ... I’m”*).

**Captioning accuracy.** In our analysis of TikTok captioning accuracy, we found that captions were largely accurate—which also reflected our participants’ experiences. We identified at least one error in 19.7% of videos (59/300), with errors in 24.0% (36/150) of Deafness and disability-related videos and in 15.3% (23/150) of general audience videos. The average word error rate (WER) among videos with at least one error was 7.9%, ranging from 0.5% to 35.7%<sup>22</sup>. Error types included word substitutions (e.g., “old on” instead of “hold on”, “rep saint of” instead of “Representative”), deletions (e.g., captioning “what’s great” as “great”), and insertions (captioning “got her dressed” as “got it her dressed”). Overall, participants reported noticing errors in TikTok captions but largely agreed with P2’s assessment that *“there’s always going to be some words that are missed or incorrect, but you basically get the overall content, and you’re able to follow.”* Errors did still impact participants’ experience, as participants skipped videos with highly inaccurate captions and stressed that when captioning *“doesn’t have as many spelling errors and word choice errors, I’ll have fewer misunderstandings”* (P5).

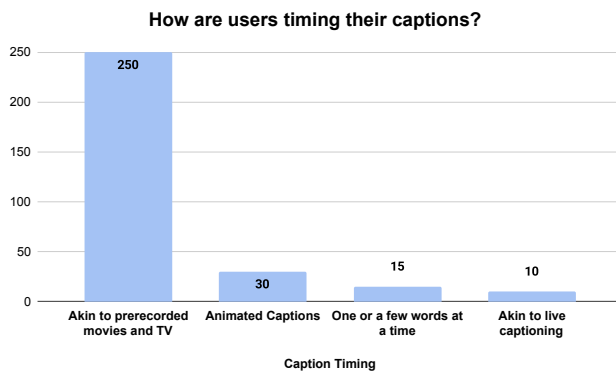
## 4.2 Participant Experiences with and Desires for User-Captioned Content

While specific facets of captioning, as we explored in Section 4.1, are crucial to video accessibility, participants also reflected broader factors shaping their use of TikTok. Here, we highlight findings on the impact of changing caption norms, perspectives on accessibility on TikTok, and desires for the future. These findings draw entirely from participant interviews.

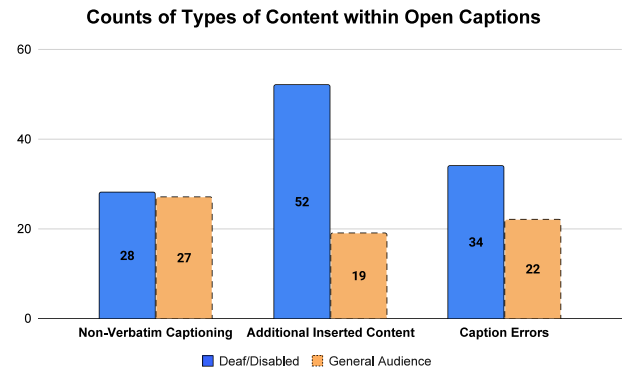
**4.2.1 Changing Captioning Norms.** Participants reflected on the impact of two changing norms in captioning: the use of automatic captioning and a shift toward open captioning.

Participants felt that automatic captioning tools produced passable TikTok captions, though they did not match the gold standard

<sup>22</sup>This analysis was conducted over 55 videos, as four were unavailable at the time of calculation



(a) Users are primarily (250/300) timing their open captions similarly to prerecorded movies and TV (a few lines at a time). Displaying captions one word at a time was less common (15/300), followed by 10/300 videos captioning similarly to live captioning, where words are shown as they are spoken and accumulate into a few lines



(b) Breakdown of frequency of occurrence for three types of content in open captions: non-verbatim captioning, inserted content, and captioning errors. Inserting additional content into open captions was most common, especially within the Deaf or Disabled dataset (52/150)

**Figure 4: Bar charts visualizing caption timing and patterns of notable content changes among general audience and Deafness and disability-related videos.**

of human captioning. Since TikTok rolled out automatic captioning in spring 2021 [74], participants reported significant increases in access: “I do feel like now with the automatic captions, almost all videos are accessible” (P9). There was, however, still a perceived drop in quality. P5 expressed his desire for “not the automated, not the kind of robotic one, but the person, the live person doing the captioning.” Many shared P6’s experience that, when viewing automatic captions, “you don’t get the full context of what they’re saying, but you kind of have, like, a broad spectrum of what they’re saying.” Participants stressed that not all automatic captioning errors have the same impact. Despite usually being able to guess at errors, P8 recounted once spending “half the video” trying to make sense of a single error—the name of the subject of the video—and concluded that “how disruptive [an error] is is not absolute.”

Closed captioning is usually considered the best practice for captioning a video, but many participants preferred the shift toward open captioning because embedding captions in a video makes it durably accessible. Some participants primarily watched TikToks off the app and, therefore, only had access to open captioned TikToks, as closed captions are an app-specific feature [42]. While some had a hard time adjusting away from closed captions, “what I’ve used most of my life” (P6), others valued that open captions were “non-turnable-offable” (P4). P9 highlighted the cultural shift that open captions represented: “[closed captioning] feels like an assistive tool instead of a complete experience, which the embedded captions do feel like.”

**4.2.2 Nuances of Platform Accessibility.** The kind of content that participants consumed significantly shaped their access needs for that video. If videos primarily contained speech, participants needed captions: “if people are just standing there talking to each other or

to the camera ... then I need to know what the words are” (P5). However, participants still reported enjoying a substantial body of uncaptioned content. P4 explained that “if it is more action, and show rather than tell” then uncaptioned videos were still worth watching (e.g., gymnastics (P1), cooking (P3), animals (P2, P4, P7)). Notably, participants had varied interests and watched a diverse range of TikTok content, emphasizing that it is important that all kinds of content are made accessible. As P3 put it, “if the video is not captioned, I’ll just be like ‘hmmmm I guess this person doesn’t care about us.’”

When participants perused TikTok individually, uncaptioned videos proved to be less of an access barrier than when engaging with TikTok socially. Participants attributed the relatively high amount of captioned content they were shown to TikTok’s algorithmically mediated nature, which they understood to “keep your preferences ... so you can use that algorithm to watch things you like” (P9). This curated view led P8 to reflect: “I think it is quite accessible ... but I think I could argue that it is inaccessible if there are videos that don’t have captions on them, I just don’t see many of them.” TikTok’s endless scroll design made it so that it was usually easy for participants to just “skip any videos without a [captioning] option” (P3), but this became complicated when viewing specific videos sent by family and friends. In P4’s experience, sharing and discussing videos is “part of the social engagement nowadays” and this interaction breaks down if the videos being shared are not captioned. P2 shared that when her siblings send her uncaptioned TikToks she’ll respond “Hello! – I’m Deaf”, to which they reply “oh, right, sorry” and then explain the content of the video.

Participants did not all feel similarly about the volume of inaccessible content on TikTok. Some felt that “TikTok videos are not like a ‘need’ thing ... it’s free entertainment” and therefore did not

take issue with the fact that not everything is accessible since “you will find something eventually” (P7). However, others recounted that their reaction to inaccessible content was “a sense of resignation” because “it’s frustrating, honestly. It means that Deaf people are falling further and further behind” (P5).

**4.2.3 Desires for the Future.** When considering what they’d like to see on TikTok in the future, participants had one overarching desire: “I would love it if every video was captioned” (P8). P1 envisioned this world: “I would love to wake up in the morning and just go, ‘oh, I can tell what’s going on.’” Toward that goal, participants considered how creators could better prioritize access, ways to integrate captioning standards and guidelines into the platform, and opportunities for customization.

Participants highlighted the ways that individual creators’ choices shape video accessibility and proposed ways to improve norms on the platform. P3 reflected that, while TikTok creators often start with some knowledge of captioning, there is room for improvement: “I think people are so much better at captioning their videos, but they’re still learning to caption it in an accessible way that is also enjoyable.” Many saw creators’ investment in captioning as a way to win their viewership: “my time, time is valuable right? And, basically, I’m going to give the reward to watch something to someone who’s invested time to make it accessible” (P2). When considering how to move toward more captioned content, P5 reflected that, rather than a technical approach, creators should “maybe just listen to us, I guess” and prioritize including Deaf and disabled viewers. Recognizing that creators are key stakeholders and that captioning is effortful, participants proposed ways that TikTok’s design could encourage and support creators in adding captions. P6 imagined adding a way to contact creators about their video accessibility, hoping that direct feedback would help creators realize “this would really help. And then it doesn’t get lost.” P7 also envisioned that TikTok could help teach creators how and why to caption videos: “whenever they are posting something, they can have like a prompt ... ‘do you wanna caption the video’ or... benefits of captioning.” Multiple participants also noted that TikTok, as a platform, could build in automatic captioning by default, making it so that if higher-quality creator-generated captions were not available, the video would retain a modicum of accessibility.

Across the board, participants noted the lack of guidelines for TikTok captions. P2 contrasted the state of most captioned media to TikTok: “when it comes to the captioning industry, there are rules, there are standards, and they know what they are. But TikTok, it’s wide open, anything goes. It’s an open source.” Participants suggested that there should be a way “to clarify some rules” (P5) for captioning in a way that would still “allow for a little creativity” (P2). For P3, this looked like building guidance into the tools creators use to make videos: “it would be really fun if everyone had a selection of captioning styles to choose from that they know would be really accessible ... [and] some technology to tell them ‘hey, your captions are overlapping this and that, let’s move them to a different place.’”

Participants also imagined that the platform would become more accessible if captions were customizable. P1 reflected that, when watching streaming television, “you can actually pick your own background and color of the [captions]—that’s really awesome.” Others noted their experience with platforms like Zoom, where they

can “drag [captions] and move them anywhere on the screen” (P3). P6 stressed that TikTok has the opportunity to not have to make captioning a “one size fits all” experience, and that customization would lead to a more accessible experience. P7 believed that being able to change the color, resize, move, or turn off captions was also key to improving the user experience of TikTok. This customizability could extend to being able to configure settings that instructed TikTok: “don’t even bother to send me things that are not, you know, captioned or whatever” (P1).

## 5 DISCUSSION

Our findings highlight relevant factors to consider when assessing how a user-generated video is captioned and point to a need for greater standardization of user-generated captions. We, therefore, discuss steps toward a captioning standard for user-generated social video, consider the future of user-driven captioning, and envision how disability justice concepts can help guide future user-generated captioning efforts.

### 5.1 Steps Toward a Captioning Standard for User-Generated Content

Having analyzed the current user-generated captioning practices in our TikTok dataset, we compare these unregulated, user-generated approaches to formal captioning standards. Participants frequently made sense of TikTok captions in relation to their understanding of standard practices, indicating that participants’ preferences for captions were strongly influenced by such standards. We therefore compare our findings on the current state of TikTok captioning with an established standard—the Described and Captioned Media Program (DCMP)—, as an exploration of what a future user-generated captioning standard could consider. Because standards vary internationally and have been shown to shape geographically-specific captioning preferences [78], we compare our findings from English-language videos and interviews with US-based participants to an American standard. While the Federal Communications Commission (FCC) is the US regulatory group that controls captioning, their guidelines are broad, focusing primarily on “accuracy, synchronicity, program completeness, and placement” [26]. However, incorporating FCC rulings and a wide body of research, the Described and Captioned Media Program (DCMP), a project by the National Association of the Deaf, has developed a comprehensive set of standards known as the Captioning Key [30]. The DCMP’s level of detail allows us to identify specific areas where the user-generated practices we observed align with or diverge from a respected standard. Their guidance is also applicable to both open and closed captions, in line with current practices on TikTok. In the following list, we compare our findings, as relevant,<sup>23</sup> to the DCMP’s major sections—text, language mechanics, presentation rate, sound effects and music, speaker identification, and special considerations.

- **Capitalization.** DCMP standards recommend mixed-case capitalization (e.g., “My dog and I played fetch.”) except to indicate shouting. While not the norm, we observed videos that used no capitalization, mixed case within words (e.g., WoOoOo), and all-caps regardless of sound volume.

<sup>23</sup>The Captioning Key is targeted toward captioning creators, so not all recommendations are relevant or information we tracked in our analysis

- **Typeface and Color.** The DCMP narrowly recommends captions use the same typeface and use white text over a translucent box. We found limited typeface variation, but a wide variety of combinations of black and white, as well as multicolored captions.
- **Caption Rate.** Per the DCMP, captions should be a minimum duration of 40 frames (slightly over one second) and a maximum of six seconds. Caption rates should also stay between 130 and 160 WPM. While we did not quantify caption duration or speed, participants reported that TikTok captions felt too fast, and we observed captions that updated with each word. Prior work finds that captions are maximally readable at 145 words per minute, but that this varies with a person's experience using captions [48].
- **Caption Placement.** Standards recommend placing captions at the bottom or, as a backup, at the top of the screen, moving captions left to right to identify speakers during dialogue. We observed captions moving across the entire screen with no clear norms for how placement can differentiate information.
- **Punctuation.** The DCMP stresses adhering to formal punctuation rules, but we observed both a lack of punctuation and creative use of punctuation.
- **Censorship.** The DCMP explicitly instructs creators to caption profanity and slang verbatim. We observed some non-verbatim captioning, and findings suggest that, on social video sharing platforms that censor videos, captioning guidelines must account for the fact that creators' choice to generate non-verbatim captions can come not from paternalism toward captioning users, but from material consequences due to platform censorship practices [79].
- **Sound Effects.** Standard captioning of sound effects does not caption all non-speech sounds but prioritizes those necessary to understand or enjoy a video. This aligned with participant expectations but not with trends in our TikTok data—very few videos captioned non-human utterances. The DCMP provides a format for captioning sound effects or describing the quality of sounds (e.g., [whispering]), but we observed a greater variety of approaches (e.g., \*whispering\*).
- **Music.** The DCMP requires that instrumental music be described only when it is essential to understanding the video but suggests that music lyrics should always be captioned. We observed that music, instrumental and lyrical alike, was rarely captioned in our datasets, but participants did not identify this as a problem. They did not want the additional cognitive load of descriptions or transcription of music that was not vital to their understanding of a video. If music was captioned, participants prioritized mood descriptions over transcription of lyrics.
- **Speaker Identification.** The DCMP suggests captions identify speakers by being placed underneath the current speaker and to identifying each speaker by name, but, in our datasets, captions were more likely to use different colors than names to differentiate speakers and varied placement both vertically and horizontally.

Ultimately, we find that any future guidelines for user-generated captioning should build from traditional standards, with key points of departure. While current standards stress formal mechanics of grammar, language, and punctuation, the looser standard we observed seemed to be appropriate for the tone of videos and was not a notable accessibility barrier. Regarding font and color, the current state of color use on TikTok often resulted in less readable captions, but participants agreed that a greater range of caption color than the DCMP's recommendation could be useful, if readability is prioritized. Although current standards recommend captioning all music, smaller screens and a different artistic role of music in user-generated social videos suggest that music should be captioned sparingly to lessen cognitive load. Our small set of participants' initial reaction highlighted that captions on TikTok are displayed too rapidly, and future work should explore both an optimal captioning rate and presentation style, taking into account the impact of varied literacy, hearing status, and experience using captions. Finally, the algorithmic censorship of videos on platforms like TikTok raises questions about verbatim captioning, and future standards may consider what kinds of non-verbatim captioning methods preserve information access without risking content removal and shadow-banning [55, 79]. Recent work by Klug et al. [55] found that TikTok creators largely use non-verbatim 'algospeak' to evade algorithmic consequences, suggesting that future user-generated captioning standards must account for the content moderation behaviors of video-hosting platforms. Our findings demonstrate a need for guidance to ensure that user-generated captions successfully extend video accessibility, and we present this comparison as a first step toward shaping future standards.

## 5.2 Toward The Future of User-Generated Captioning

The videos in our dataset are representative of a new era of considerations for captioning: they are open-captioned by users engaged in internet culture. This poses new considerations for captioning design and research, namely how to engage video creators, who to study as captioning users, and how to systematically study open captions.

Traditional captioning tool design either assumes captions will be generated by professional CART captioners (e.g., [46, 53, 59]), or by automatic speech recognition-based tools (e.g., [23, 70]), and therefore does not consider the needs of non-expert captioners. Video creators are fundamental to the existence of user-generated captioning, and our findings reveal many avenues for future change that require significant effort from video creators. Future platform design should consider ways to both incentivize and enforce high-quality captioning, and future work needs to engage video creators in the design of those systems.

Captions have traditionally been studied as a tool used by DHH people (e.g., [16, 28, 59]), but recent research has emphasized that other disability communities, particularly neurodivergent people, also use captions to access audio/video content [85]. Correlating assistive technology use with a single disability group thus misses the perspectives of these other potential users [68]. Further, even within Deaf and hard of hearing captioning users, preferences and experiences can differ [60, 90]. To account for these varied users

and experiences, we explicitly recruited “captioning users” broadly rather than focusing on a specific group such as DHH participants. However, all but one of our participants identified as DHH, which means that we were unable to explore tensions among the needs of different groups of captioning users—an important direction for future work. Analogously, past work has found that users of alt text, another user-created digital access tool, have a variety of preferences [66, 89]. Researchers and designers have begun to propose approaches to alt text provision that meet varied needs, namely customization. In a similar vein, understanding and including the needs of all people who use captions to meet an access need is crucial to ensuring an inclusive future of captioning design.

Finally, the shift toward open captioning produces a new set of considerations when assessing captioning quality. Recently, HCI captioning researchers have emphasized the importance of metrics to understand and improve caption accuracy [50] and placement [9]. The set of features we analyzed (see Table 2) could serve as a step toward a structured analytical tool for understanding the quality of open captions. Such a tool could support future researchers in assessing factors beyond accuracy, holistically encompassing elements of audio coverage, design, and captioning content, which are necessary to consider when engaging with open-captioned videos. These features could also be useful in creating future tools to guide non-expert caption creators in making considered decisions when generating new captions or understanding the state of their past content.

### 5.3 Disability Justice and TikTok Access

Accessibility legislation and research overwhelmingly focus on access to critical or educational information, often to the exclusion of entertainment or content deemed less important. Within HCI captioning literature, research overwhelmingly focuses on access to education (e.g., [53, 59]), work (e.g., [14, 69]), or informative media (e.g., news [28], education [16]). However, while a few participants used TikTok for informative purposes, most recounted enjoying watching silly pet, cooking, and trending dance videos. In fact, some participants wrestled with the idea that something that is “*not like a ‘need’ thing*” (P7) ought to be accessible. We argue that ensuring accessibility to content, even when it does not fulfill a specific need, is essential—all people deserve access to idle entertainment and the ability to participate in the “*social engagement*” (P4) of sharing and discussing silly videos.

For user-generated content to become accessible content, we argue that creators must embrace principles of disability justice, particularly collective access and interdependence. The disability justice principles of collective access - accessibility is a group, not individual responsibility - and interdependence - that we all rely upon each other to navigate the world - articulate a world where everyone is responsible for considering how to extend access to others [45]. Prior work has often been motivated by the idea that not all online videos will be well-captioned (e.g., [16, 84]). We argue that by adopting a lens of collective access and interdependence, we can imagine a world where high-quality captioning is seen as inherent to user-generated video content and focus future efforts on building tools that help realize that world.

## 6 LIMITATIONS

Our study has a few key limitations. First, although we reached saturation while analyzing 300 videos, this represents a fraction of the videos uploaded to TikTok every minute. The type of mixed methods analysis we conducted does not scale indefinitely, and future quantitative analyses of TikTok captioning at scale could complement this work. Second, we scoped our dataset to English language videos because of our research team’s fluency in English. All participants were also based in the US. Future work should examine how captions and caption users’ perspectives vary in non-English language and international contexts. Next, we focused on interviewing individuals who already use TikTok and therefore cannot address whether viewers who need captions to access videos find TikTok to be accessible overall. Our participants considered TikTok to be accessible enough to be enjoyable, but we cannot speculate whether this perspective holds universally. Further, we intentionally defined our recruitment criteria based on use of captions to meet an access need, rather than a specific d/Deaf or disability identity. However, all but one participant identified as DHH. Future research should seek to have greater participant diversity. Finally, nine participants is a small sample. Future research, particularly work exploring a standard for user-generated captioning, should seek to validate our findings with a much larger participant pool.

## 7 CONCLUSION

As the world’s most downloaded app in 2020, [73] and a massive repository of user-generated video content, TikTok provides an exciting opportunity to understand current trends in user-generated captioning and explore how those captions impact the many caption users viewing TikTok daily. We conducted a content analysis of 300 TikTok videos, evenly distributed between general audience and Deafness and disability datasets, and interviewed nine frequent TikTok viewers who rely on captions. Our findings reveal that current TikTok captioning practices facilitate access but could be improved, perhaps with the aid of a user-generated content-specific captioning standard. This work contributes the first empirical understanding of the state of captioning on TikTok, and provides approaches to advancing toward a world with universal captioning for user-generated content.

## ACKNOWLEDGMENTS

We thank our colleagues Aashaka Desai and Steven Goodman for their support and input over the course of this study. Our work was made possible by captioners, interpreters, and the University of Washington’s Deaf and hard of hearing services coordinator, Dimitri Azadi. We thank our participants for their invaluable insights. This work was partly supported by the National Science Foundation under Grant No. IIS-1763199, the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-2140004, and by the University of Washington’s CREATE center.

## REFERENCES

- [1] 2015. Closed Captioning Quality Report and Order, Declaratory Ruling, FN-PRM. <https://www.fcc.gov/document/closed-captioning-quality-report-and-order-declaratory-ruling-fnprm>
- [2] 2021. Closed Captioning of Internet Video Programming. <https://www.fcc.gov/consumers/guides/captioning-internet-video-programming>



- [3] 2022. <https://www.3playmedia.com/learn/accessibility/accessibility-laws/>
- [4] 2022. <https://www.bbc.com.uk/accessibility/forproducts/guides/subtitles/>
- [5] Oliver Alonzo, Hijung Valentina Shin, and Dingzeyu Li. 2022. Beyond Subtitles: Captioning and Visualizing Non-speech Sounds to Improve Accessibility of User-Generated Videos. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '22)*. Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3517428.3544808>
- [6] Meryl Alper, Jessica Sage Rauchberg, Ellen Simpson, Josh Guberman, and Sarah Feinberg. 2023. TikTok as algorithmically mediated biographical illumination: Autism, self-discovery, and platformed diagnosis on# autistoktok. *New Media & Society* (2023), 14614448231193091.
- [7] Akhter Al Amin. 2020. Audio-Visual Caption Evaluation Metric for People who are Deaf and Hard of Hearing. *Articles* (July 2020). <https://scholarworks.rut.edu/article/1964>
- [8] Akhter Al Amin, Saad Hassan, and Matt Huenerfauth. 2021. Caption-occlusion severity judgments across live-television genres from deaf and hard-of-hearing viewers. In *Proceedings of the 18th International Web for All Conference (W4A '21)*. Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3430263.3452429>
- [9] Akhter Al Amin, Saad Hassan, Sooyeon Lee, and Matt Huenerfauth. 2022. Watch It, Don't Imagine It: Creating a Better Caption-Occlusion Metric by Collecting More Ecologically Valid Judgments from DHH Viewers. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3491102.3517681>
- [10] Akhter Al Amin, Joseph Mendis, Raja Kushalnagar, Christian Vogler, and Matt Huenerfauth. 2023. Who is Speaking: Unpacking In-Text Speaker Identification Preference of Viewers Who Are Deaf and Hard of Hearing While Watching Live Captioned Television Program. In *Proceedings of the 20th International Web for All Conference (<conf-loc>, <city>Austin</city>, <state>TX</state>, <country>USA</country>, </conf-loc>)* (W4A '23). Association for Computing Machinery, New York, NY, USA, 44–53. <https://doi.org/10.1145/3587281.3587286>
- [11] Corey H Basch, Zoe Meleo-Erwin, Joseph Fera, Christie Jaime, and Charles E Basch. 2021. A global pandemic in the time of viral memes: COVID-19 vaccine misinformation and disinformation on TikTok. *Human vaccines & immunotherapeutics* 17, 8 (2021), 2373–2377.
- [12] Collin Matthew Belt. 2013. American Sign Language is Not English on the Hands. <https://www.lifeprint.com/asl101/topics/history8.htm>
- [13] Larwan Berke, Khaled Albusays, Matthew Seita, and Matt Huenerfauth. 2019. Preferred Appearance of Captions Generated by Automatic Speech Recognition for Deaf and Hard-of-Hearing Viewers. (2019). <https://doi.org/10.1145/3290607.3312921> ISBN: 9781450359719.
- [14] Larwan Berke, Christopher Caulfield, and Matt Huenerfauth. 2017. Deaf and Hard-of-Hearing Perspectives on Imperfect Automatic Speech Recognition for Captioning One-on-One Meetings. (2017). <https://doi.org/10.1145/3132525.3132541> ISBN: 9781450349260.
- [15] Larwan Berke, Sushant Kafle, and Matt Huenerfauth. 2018. Methods for Evaluation of Imperfect Captioning Tools by Deaf or Hard-of-Hearing Users at Different Reading Literacy Levels. (2018). <https://doi.org/10.1145/3173574.3173665> ISBN: 9781450356206.
- [16] Larwan Berke, Matthew Seita, and Matt Huenerfauth. 2020. Deaf and hard-of-hearing users' prioritization of genres of online video content requiring accurate captions. In *Proceedings of the 17th International Web for All Conference (W4A '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3371300.3383337>
- [17] Aparajita Bhandari and Sara Bimo. 2022. Why's everyone on TikTok now? the algorithmized self and the future of self-making on social media. *Social Media + Society* 8, 1 (Mar 2022). <https://doi.org/10.1177/20563051221086241>
- [18] Sydney Bradley and Dan Whateley. 2023. TikTok has surged past Instagram in time spent on the platform, but lags in ad dollars. <https://www.businessinsider.com/tiktok-surpasses-instagram-time-spent-us-adults-emarketer-8-2023>
- [19] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. <https://doi.org/10.1191/1478088706qp0630a> Publisher: Routledge \_eprint: <https://www.tandfonline.com/doi/pdf/10.1191/1478088706qp0630a>
- [20] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4 (Aug. 2019), 589–597. <https://doi.org/10.1080/2159676X.2019.1628806> Publisher: Routledge \_eprint: <https://doi.org/10.1080/2159676X.2019.1628806>
- [21] Andy Brown, Rhia Jones, Mike Crabb, James Sandford, Matthew Brooks, Mike Armstrong, and Caroline Jay. 2015. Dynamic Subtitles: The User Experience. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video (TVX '15)*. Association for Computing Machinery, New York, NY, USA, 103–112. <https://doi.org/10.1145/2745197.2745204>
- [22] Janine Butler. 2020. The Visual Experience of Accessing Captioned Television and Digital Videos. *Television & New Media* 21, 7 (Nov. 2020), 679–696. <https://doi.org/10.1177/1527476418824805> Publisher: SAGE Publications.
- [23] Janine Butler, Brian Trager, and Byron Behm. 2019. Exploration of Automatic Speech Recognition for Deaf and Hard of Hearing Students in Higher Education Classes. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '19)*. Association for Computing Machinery, Pittsburgh, PA, USA, 32–42. <https://doi.org/10.1145/3308561.3353772>
- [24] Sebastian Cahill. 2023. From tiktok to netflix, gen Z is embracing closed captions. we asked them why. <https://www.insider.com/generation-z-subtitles-closed-captioning-millennials-no-hearing-loss-2023-8>
- [25] Karen Collins and Peter J. Taillon. 2012. Visualized sound effect icons for improved multimedia accessibility: A pilot study. *Entertainment Computing* 3, 1 (Jan. 2012), 11–17. <https://doi.org/10.1016/j.entcom.2011.09.002>
- [26] Consumer and Governmental Affairs. 2014. *Closed Captioning of Video Programming; Telecommunications for the Deaf and Hard of Hearing, Inc.* Number FCC 14-12. <https://www.fcc.gov/document/closed-captioning-quality-report-and-order-declaratory-ruling-fmpm>
- [27] Brian Contreras. 2021. Why captions are everywhere on TikTok: “glasses for your ears”. <https://www.latimes.com/business/technology/story/2021-09-23/subtitles-once-a-barrier-to-foreign-film-have-found-a-home-on-tiktok>
- [28] Michael Crabb, Rhianne Jones, Mike Armstrong, and Chris J. Hughes. 2015. Online News Videos: The UX of Subtitle Position. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS '15)*. Association for Computing Machinery, New York, NY, USA, 215–222. <https://doi.org/10.1145/2700648.2809866>
- [29] Maitraye Das, John Tang, Kathryn E Ringland, and Anne Marie Piper. 2021. Towards accessible remote work: Understanding work-from-home practices of neurodivergent professionals. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–30.
- [30] DCMF. 2023. Captioning Key. <https://dcmp.org/learn/captioningkey>
- [31] Michael Ann DeVito. 2022. How Transfeminine TikTok Creators Navigate the Algorithmic Trap of Visibility Via Folk Theorization. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 380:1–380:31. <https://doi.org/10.1145/3555105>
- [32] DO-IT. 2021. What is the difference between open and closed captioning? <https://www.washington.edu/doit/what-difference-between-open-and-closed-captioning>
- [33] Katie Dupere. 2021. Deaf youtubers call out crappy captions with Hashtag-nomorecaptions movement. <https://mashable.com/2016/11/11/youtube-closed-captioning-nomorecaptions/>
- [34] Jared Duval, Ferran Altarriba Bertran, Siying Chen, Melissa Chu, Divya Subramonian, Austin Wang, Geoffrey Xiang, Sri Kurniawan, and Katherine Isbister. 2021. Chasing Play on TikTok from Populations with Disabilities to Inspire Playful and Inclusive Technology Design. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–15. <https://doi.org/10.1145/3411764.3445303>
- [35] Tessa Eagle and Kathryn E. Ringland. 2023. “You Can’t Possibly Have ADHD”: Exploring Validation and Tensions around Diagnosis within Un-bounded ADHD Social Media Communities. In *The 25th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '23)*. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3597638.3608400>
- [36] Elizabeth Ellcessor. 2012. Captions On, Off, on TV, Online: Accessibility and Search Engine Optimization in Online Closed Captioning. *Television & New Media* 13, 4 (July 2012), 329–352. <https://doi.org/10.1177/1527476411425251> Publisher: SAGE Publications.
- [37] Moa Eriksson Krutrök. 2021. Algorithmic Closeness in Mourning: Vernaculars of the Hashtag #grief on TikTok. *Social Media + Society* 7, 3 (Jul 2021), 20563051211042396. <https://doi.org/10.1177/20563051211042396>
- [38] Jayden Galamgam and Justin L. Jia. 2021. “Accutane Check”: insights into youth sentiment toward isotretinoin from a TikTok trend. *Pediatric Dermatology* 38, 4 (2021), 980–981.
- [39] Abraham Glasser, Kesavan Kushalnagar, and Raja Kushalnagar. 2019. Deaf, Hard of Hearing, and Hearing Perspectives on using Automatic Speech Recognition in Conversation. *arXiv:1909.01176 [cs]* (Sept. 2019). <https://doi.org/10.1145/3132525.3134781> arXiv: 1909.01176.
- [40] Michael Gower, Brent Shiver, Charu Pandhi, and Shari Trewin. 2018. Leveraging Pauses to Improve Video Captions. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '18)*. Association for Computing Machinery, Galway, Ireland, 414–416. <https://doi.org/10.1145/3234695.3241023>
- [41] Shannon S. C. Herrick, Laura Hallward, and Lindsay R. Duncan. 2021. “This is just how I cope”: An inductive thematic analysis of eating disorder recovery content created and shared on TikTok using EDrecovery. *The International Journal of Eating Disorders* 54, 4 (Apr 2021), 516–526. <https://doi.org/10.1002/eat.23463>
- [42] Stephanie Hind. 2021. Introducing auto captions. <https://newsroom.tiktok.com/en-us/introducing-auto-captions>

- [43] Richang Hong, Meng Wang, Mengdi Xu, Shuicheng Yan, and Tat-Seng Chua. 2010. Dynamic captioning: video accessibility enhancement for hearing impairment. In *Proceedings of the 18th ACM international conference on Multimedia (MM '10)*. Association for Computing Machinery, Firenze, Italy, 421–430. <https://doi.org/10.1145/1873951.1874013>
- [44] Yongtao Hu, Jan Kautz, Yizhou Yu, and Wenping Wang. 2015. Speaker-Following Video Subtitles. *ACM Transactions on Multimedia Computing, Communications, and Applications* 11, 2 (Jan. 2015), 1–17. <https://doi.org/10.1145/2632111>
- [45] Sins Invalid. 2019. *Skin Tooth and Bone: The Basis of Movement is Our People, a Disability Justice Primer* (2nd ed.). Sins Invalid.
- [46] Dhruv Jain, Leah Findlater, Jamie Gilkeson, Benjamin Holland, Ramani Duraiswami, Dmitry Zotkin, Christian Vogler, and Jon E. Froehlich. 2015. Head-Mounted Display Visualizations to Support Sound Awareness for the Deaf and Hard of Hearing. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Association for Computing Machinery, Seoul, Republic of Korea, 241–250. <https://doi.org/10.1145/2702123.2702393>
- [47] Cordilia James. 2022. Why do all these 20-somethings have closed captions turned on? <https://www.wsj.com/articles/why-do-all-these-20-somethings-have-closed-captions-turned-on-11663386473>
- [48] Carl Jensema. 1998. Viewer Reaction to Different Television Captioning Speeds. *American Annals of the Deaf* 143, 4 (1998), 318–324. <https://doi.org/10.1353/aad.2012.0073>
- [49] C. Jensema, R. McCann, and S. Ramsey. 1996. Closed-captioned television presentation speed and vocabulary. *American Annals of the Deaf* 141, 4 (Oct. 1996), 284–292. <https://doi.org/10.1353/aad.2012.0377>
- [50] Sushant Kafle and Matt Huenerfauth. 2017. Evaluating the Usability of Automatically Generated Captions for People who are Deaf or Hard of Hearing. (2017). <https://doi.org/10.1145/3132525.3132542> ISBN: 9781450349260.
- [51] Sushant Kafle, Peter Yeung, and Matt Huenerfauth. 2019. Evaluating the Benefit of Highlighting Key Words in Captions for People who are Deaf or Hard of Hearing. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, Pittsburgh PA USA, 43–55. <https://doi.org/10.1145/3308561.3353781>
- [52] Nadia Karizat, Dan Delmonaco, Motahhare Eslami, and Nazanin Andalibi. 2021. Algorithmic Folk Theories and Identity: How TikTok Users Co- Produce Knowledge of Identity and Engage in Algorithmic Resistance. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 305:1–305:44. <https://doi.org/10.1145/3476046>
- [53] Saba Kwas, George Karalis, Tzu Wen, and Richard E Ladner. 2016. Improving Real-Time Captioning Experiences for Deaf and Hard of Hearing Students. (2016). <https://doi.org/10.1145/2982142.2982164> ISBN: 9781450341240.
- [54] Daniel Klug, Yiluo Qin, Morgan Evans, and Geoff Kaufman. 2021. Trick and Please. A Mixed-Method Study On User Assumptions About the TikTok Algorithm. In *13th ACM Web Science Conference 2021 (WebSci '21)*. Association for Computing Machinery, New York, NY, USA, 84–92. <https://doi.org/10.1145/3447535.3462512>
- [55] Daniel Klug, Ella Steen, and Kathryn Yurechko. 2023. How Algorithm Awareness Impacts Algospreak Use on TikTok. In *Companion Proceedings of the ACM Web Conference 2023 (Austin, TX, USA) (WWW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 234–237. <https://doi.org/10.1145/3543873.3587355>
- [56] Wenwen Kong, Shijie Song, Yuxiang Chris Zhao, Qinghua Zhu, and Ling Sha. 2021. TikTok as a Health Information Source: Assessment of the Quality of Information in Diabetes-Related Videos. *Journal of Medical Internet Research* 23, 9 (Sep 2021), e30409. <https://doi.org/10.2196/30409>
- [57] Kuno Kurzhals, Fabian Göbel, Katrin Angerbauer, Michael Sedlmair, and Martin Raubal. 2020. A View on the Viewer: Gaze-Adaptive Captions for Videos. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3313831.3376266>
- [58] Raja Kushalnagar and Kesavan Kushalnagar. 2018. Subtitleformatter: Making subtitles easier to read for deaf and hard of hearing viewers on personal devices. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 10896 LNCS. Springer Verlag, 211–219. [https://doi.org/10.1007/978-3-319-94277-3\\_35](https://doi.org/10.1007/978-3-319-94277-3_35) ISSN: 16113349.
- [59] Raja S. Kushalnagar, Gary W. Behm, Aaron W. Kelstone, and Shareef Ali. 2015. Tracked Speech-To-Text Display: Enhancing Accessibility and Readability of Real-Time Speech-To-Text. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS '15)*. Association for Computing Machinery, Lisbon, Portugal, 223–230. <https://doi.org/10.1145/2700648.2809843>
- [60] Raja S. Kushalnagar, Gary W. Behm, Kevin T. Wolfe, Peter Yeung, Becca Dingman, Shareef Sayel Ali, Abraham Glasser, and Claire Elizabeth Ryan. 2018. RTTD-ID: Tracked Captions with Multiple Speakers for Deaf Students. <https://peer.asee.org/rttd-id-tracked-captions-with-multiple-speakers-for-deaf-students>
- [61] Daniel G. Lee, Deborah I. Fels, and John Patrick Udo. 2007. Emotive captioning. *Computers in Entertainment* 5, 2 (April 2007), 11. <https://doi.org/10.1145/1279540.1279551>
- [62] Rachel Lerman. 2021. Social media has upped its accessibility game. But deaf creators say it has a long way to go. *Washington Post* (April 2021). <https://www.washingtonpost.com/technology/2021/03/15/social-media-accessibility-captions/>
- [63] Franklin Mingzhe Li, Cheng Lu, Zhicong Lu, Patrick Carrington, and Khai N. Truong. 2022. An Exploration of Captioning Practices and Challenges of Individual Content Creators on YouTube for People with Hearing Impairments. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (April 2022), 75:1–75:26. <https://doi.org/10.1145/3512922>
- [64] Chen Ling, Krishna P Gummadi, and Savvas Zannettou. 2023. "Learn the Facts About COVID-19": Analyzing the Use of Warning Labels on TikTok Videos. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 17. 554–565.
- [65] Xingyu "Bruce" Liu, Ruolin Wang, Dingzeyu Li, Xiang Anthony Chen, and Amy Pavel. 2022. CrossA11y: Identifying Video Accessibility Issues via Cross-modal Grounding. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology (UIST '22)*. Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3526113.3545703>
- [66] Kelly Mack, Edward Cutrell, Bongshin Lee, and Meredith Ringel Morris. 2021. Designing Tools for High-Quality Alt Text Authoring. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility (Virtual Event, USA) (ASSETS '21)*. Association for Computing Machinery, New York, NY, USA, Article 23, 14 pages. <https://doi.org/10.1145/3441852.3471207>
- [67] Kelly Mack, Emma McDonnell, Dhruv Jain, Lucy Lu Wang, Jon E. Froehlich, and Leah Findlater. 2021. What Do We Mean by "Accessibility Research"? A Literature Survey of Accessibility Papers in CHI and ASSETS from 1994 to 2019. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Yokohama, Japan. <https://doi.org/10.1145/3411764.3445412>
- [68] Kelly Mack, Emma J. McDonnell, Leah Findlater, and Heather D. Evans. 2022. Chronically Under-Addressed: Considerations for HCI Accessibility Practice with Chronically Ill People. In *The 24th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, Athens Greece, 1–15. <https://doi.org/10.1145/3517428.3544803>
- [69] James R. Mallory, Michael Stinson, Lisa Elliot, and Donna Easton. 2017. Personal Perspectives on Using Automatic Speech Recognition to Facilitate Communication between Deaf Students and Hearing Customers. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, Baltimore Maryland USA, 419–421. <https://doi.org/10.1145/3132525.3134779>
- [70] Emma J. McDonnell, Soo Hyun Moon, Lucy Jiang, Steven M. Goodman, Raja Kushalnagar, Jon E. Froehlich, and Leah Findlater. 2023. "Easier or Harder, Depending on Who the Hearing Person Is": Codesigning Videoconferencing Tools for Small Groups with Mixed Hearing Status. In *The ACM CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg, Germany.
- [71] Ashlee Milton, Leah Ajmani, Michael Ann DeVito, and Stevie Chancellor. 2023. "I See Me Here": Mental Health Content, Community, and Algorithmic Curation on TikTok. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–17. <https://doi.org/10.1145/3544548.3581489>
- [72] NAD. 2023. Community and culture – frequently asked questions. <https://www.nad.org/resources/american-sign-language/community-and-culture-frequently-asked-questions/>
- [73] BBC News. 2021. TikTok named as the most downloaded app of 2020. *BBC News* (Aug 2021). <https://www.bbc.com/news/business-58155103>
- [74] Marije Nouwen and Mathilde Hermine Christine Marie Ghislaine Duflos. 2021. TikTok as a data gathering space: the case of grandchildren and grandparents during the COVID-19 pandemic. In *Interaction Design and Children (IDC '21)*. Association for Computing Machinery, New York, NY, USA, 498–502. <https://doi.org/10.1145/3459990.3465201>
- [75] Kotaro Oomori, Akihisa Shitara, Tatsuya Minagawa, Sayan Sarcar, and Yoichi Ochiai. 2020. A Preliminary Study on Understanding Voice-only Online Meetings Using Emoji-based Captioning for Deaf or Hard of Hearing Users. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '20)*. Association for Computing Machinery, New York, NY, USA, 1–4. <https://doi.org/10.1145/3373625.3418032>
- [76] Andrew D. Ouzts, Nicole E. Snell, Prabhudh Maini, and Andrew T. Duchowski. 2013. Determining optimal caption placement using eye tracking. In *Proceedings of the 31st ACM international conference on Design of communication (SIGDOC '13)*. Association for Computing Machinery, New York, NY, USA, 189–190. <https://doi.org/10.1145/2507065.2507100>
- [77] Yi-Hao Peng, Ming-Wei Hsi, Paul Taele, Ting-Yu Lin, Po-En Lai, Leon Hsu, Tzu-chuan Chen, Te-Yen Wu, Yu-An Chen, Hsien-Hui Tang, and Mike Y. Chen. 2018. SpeechBubbles: Enhancing Captioning Experiences for Deaf and Hard-of-Hearing People in Group Conversations. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, Montreal QC, Canada, 1–10. <https://doi.org/10.1145/3173574.3173867>

- [78] Raisa Rashid, Quoc Vy, Richard Hunt, and Deborah Fels. 2008. Dancing with Words: Using Animated Text for Captioning. *Int. J. Hum. Comput. Interaction* 24 (June 2008), 505–519. <https://doi.org/10.1080/10447310802142342>
- [79] Jessica Sage Rauchberg. 2022. # Shadowbanned: Queer, Trans, and Disabled creator responses to algorithmic oppression on TikTok. In *LGBTQ Digital Cultures*. Routledge, 196–209.
- [80] April Rubin. 2023. Why young people can't get enough of subtitles - axios. <https://www.axios.com/2023/08/20/gen-z-millennials-tv-movies-subtitles>
- [81] Alex M. Russell, Robert E. Davis, Juanybeth M. Ortega, Jason B. Colditz, Brian Primack, and Adam E. Barry. 2021. Alcohol: Portrayals of Alcohol in Top Videos on TikTok. *Journal of Studies on Alcohol and Drugs* 82, 5 (Sep 2021), 615–622. <https://doi.org/10.15288/jsad.2021.82.615>
- [82] Anastasia Schaadhardt, Yue Fu, Cory Gennari Pratt, and Wanda Pratt. 2023. "Laughing so I don't cry": How TikTok users employ humor and compassion to connect around psychiatric hospitalization. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [83] Alina Secară. 2011. R U ready 4 new subtitles? Investigating the potential of social translation practices and creative spellings. *Linguistica Antverpiensia, New Series – Themes in Translation Studies* 10 (2011). <https://doi.org/10.52034/lanstts.v10i.282>
- [84] Brent N. Shiver and Rosalee J. Wolfe. 2015. Evaluating Alternatives for Better Deaf Accessibility to Selected Web-Based Multimedia. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility - ASSETS '15*. ACM Press, Lisbon, Portugal, 231–238. <https://doi.org/10.1145/2700648.2809857>
- [85] Ellen Simpson, Samantha Dalal, and Bryan Semaan. 2023. "Hey, Can You Add Captions?": The Critical Infrastructuring Practices of Neurodiverse People on TikTok. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (April 2023), 1–27. <https://doi.org/10.1145/3579490>
- [86] Ellen Simpson, Andrew Hamann, and Bryan Semaan. 2022. How to Tame "Your" Algorithm: LGBTQ+ Users' Domestication of TikTok. *Proceedings of the ACM on Human-Computer Interaction* 6, GROUP (Jan. 2022), 22:1–22:27. <https://doi.org/10.1145/3492841>
- [87] Ellen Simpson and Bryan Semaan. 2021. For You, or For "You"? Everyday LGBTQ+ Encounters with TikTok. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (Jan 2021), 252:1–252:34. <https://doi.org/10.1145/3432951>
- [88] Lauren Southwick, Sharath C Guntuku, Elissa V Klinger, Emily Seltzer, Haley J McCalpin, and Raina M Merchant. 2021. Characterizing COVID-19 content posted to TikTok: public sentiment and response during the first phase of the COVID-19 pandemic. *Journal of Adolescent Health* 69, 2 (2021), 234–241.
- [89] Abigale Stangl, Nitin Verma, Kenneth R. Fleischmann, Meredith Ringel Morris, and Danna Gurari. 2021. Going Beyond One-Size-Fits-All Image Descriptions to Satisfy the Information Wants of People Who Are Blind or Have Low Vision. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility (Virtual Event, USA) (ASSETS '21)*. Association for Computing Machinery, New York, NY, USA, Article 16, 15 pages. <https://doi.org/10.1145/3441852.3471233>
- [90] Agnieszka Szarkowska, Izabela Krejtz, Zuzanna Klyszejko, and Anna Wiczorek. 2011. Verbatim, Standard, or Edited?: Reading Patterns of Different Captioning Styles Among Deaf, Hard of Hearing, and Hearing Viewers. *American Annals of the Deaf* 156, 4 (Nov. 2011), 363–378. <https://doi.org/10.1353/aad.2011.0039> Publisher: Gallaudet University Press.
- [91] TikTok. 2021. Creating Longer Videos. <https://www.tiktok.com/creators/creator-portal/en-us/tiktok-content-strategy/creating-videos-longer-than-one-minute/>
- [92] TikTok. 2023. Terms of service. <https://www.tiktok.com/legal/page/us/terms-of-service/en>
- [93] Quoc V. Vy and Deborah I. Fels. 2010. Using Placement and Name for Speaker Identification in Captioning. In *Computers Helping People with Special Needs (Lecture Notes in Computer Science)*, Klaus Miesenberger, Joachim Klaus, Wolfgang Zagler, and Arthur Karshmer (Eds.). Springer, Berlin, Heidelberg, 247–254. [https://doi.org/10.1007/978-3-642-14097-6\\_40](https://doi.org/10.1007/978-3-642-14097-6_40)
- [94] James M. Waller and Raja S. Kushalnagar. 2016. Evaluation of Automatic Caption Segmentation. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '16)*. Association for Computing Machinery, New York, NY, USA, 331–332. <https://doi.org/10.1145/2982142.2982205>
- [95] Kevin Westcott, Jana Arbanas, Chris Arkenberg, Brooke Auxier, Jeff Loucks, and Kevin Downs. 2023. 2023 digital media trends Immersed and connected. <https://www2.deloitte.com/x/en/insights/industry/technology/media-industry-trends-2023.html>
- [96] Anthony Yeung, Enoch Ng, and Elia Abi-Jaoude. 2022. TikTok and Attention-Deficit/Hyperactivity Disorder: A Cross-Sectional Study of Social Media Content Quality. *The Canadian Journal of Psychiatry* (Feb 2022), 07067437221082854. <https://doi.org/10.1177/07067437221082854>
- [97] Sean Zdenek. 2015. *Reading Sounds: Closed-Captioned Media and Popular Culture*. University of Chicago Press. Google-Books-ID: IY6wCgAAQBAJ.
- [98] Sean Zdenek. 2018. Designing Captions: Disruptive Experiments with Typography, Color, Icons, and Effects. 23.1 (Aug. 2018). <https://kairos.technorhetoric.net/23.1/topoi/zdenek/introduction.html> Publisher: Kairos: A Journal of Rhetoric, Technology, and Pedagogy.
- [99] Sean Zdenek. 2018. Designing Captions: Disruptive Experiments with Typography, Color, Icons, and Effects. 23.1 (Aug. 2018). <https://kairos.technorhetoric.net/23.1/topoi/zdenek/index.html> Publisher: Kairos: A Journal of Rhetoric, Technology, and Pedagogy.
- [100] David X. Zheng, Anne Y. Ning, Melissa A. Levoska, Laura Xiang, Christina Wong, and Jeffrey F. Scott. 2021. Acne and social media: A cross-sectional study of content quality on TikTok. *Pediatric Dermatology* 38, 1 (Jan 2021), 336–338. <https://doi.org/10.1111/pde.14471>
- [101] Diana Zulli and David James Zulli. 2020. Extending the internet meme: Conceptualizing technological mimesis and imitation publics on the TikTok platform. *New Media & Society* 24, 8 (2020), 1872–1890. <https://doi.org/10.1177/1461444820983603>