"I Just Didn't Notice It:" Experiences with Misinformation Warnings on Social Media amongst Users Who Are Low Vision or Blind

Filipo Sharevski
DePaul University
Chicago, IL, USA
fsharevs@cdm.depaul.edu

Aziz Zeidieh
University of Illinois at Urbana-Champaign
Champaign, IL, USA
azeidi2@illinois.edu

ABSTRACT

Dealing with misinformation on social media is a complex affair as platforms have to continuously decide whether and how to moderate falsehoods and misleading content. The options available are either hard moderation i.e. content and account removal or soft moderation i.e. substantiate false or misleading posts with misinformation warning labels. These warning labels are implemented as *visual frictions* with the intention to interrupt the user's immersive experience and "nudge" them towards a better truth discernment. The choice of visual friction poses the question whether these warning labels are accessible for users who are low vision or blind. From the first accounts of 29 such users in our study, we learned that this is not the case. Excluded as such, the misinformation warning labels we tested on three platforms – Facebook, YouTube, and TikTok – did not help 72.4% of the visually impaired participants towards a better truth discernment. Our participants, therefore, provided useful and actionable recommendations for *inclusive* design of misinformation warnings that could meaningfully help the overall effort for curbing falsehoods and misleading statements.

CCS CONCEPTS

• **Security and privacy → Social aspects of security and privacy**; **Usability in security and privacy**; • **Human-centered computing → Social networks**.

KEYWORDS

accessibility, blind, low vision, visually impaired, social media, misinformation, soft moderation, warning labels, truth discernment

ACM Reference Format:
Filipo Sharevski and Aziz Zeidieh. 2021. "I Just Didn't Notice It:" Experiences with Misinformation Warnings on Social Media amongst Users Who Are Low Vision or Blind. In *New Security Paradigms Workshop (NSPW '23), September 18-21, 2023, 2023,Segovia, Spain.* ACM, New York, NY, USA, 16 pages. https://doi.org/DOI

NSPW '23, September 18-21, 2023, Segovia, Spain

ACM ISBN ISBN.
https://doi.org/DOI

1 INTRODUCTION

Misinformation is popular both as a content on social media as well as a topic of inquiry among academic researchers [65, 76, 97]. Falsehoods and misleading statements provide unique experiences for users on all platforms, from the usual mainstream places to alternative fringe communities online [74]. Researchers do work to detect and remove misinformation on a platform level [81, 96, 109], but also to implement and improve misinformation interventions when warning about a potentially misleading of false content on a user level [75, 85].

The scientific effort against misinformation is driven towards preventing *harmful* experiences of misinformation exposure that range from emotional distress, opinion manipulation, to erosion of interpersonal relationships among people [37]. While the automatic detection of misleading or false statements helps limit the circulation of harmful content on social media (or at least on the mainstream platforms), the misinformation warnings do allow for exposure and leave the people themselves to decide if and how they will discern the truthfulness of this content. Psychology, political science, and behavioural researchers, concerned with the effectiveness of these warnings for better truth discernment and informed decision-making, have learned that misinformation interventions received a mixed response from social media users as they often find them uninformative [15], intrusive [70], and counterproductive [21, 88]. Security and human-interaction researchers, concerned with the usability and the deception-prevention ability of these warnings, appended this knowledge with evidence that the misinformation interventions are indistinguishable from the platforms' aesthetics [13], lack context [75], and confuse users without eliminating the deceptive effect [46, 72].

Actionable as is, the research about warning social media users about potentially misleading or false content has failed to *include* the needs of all users on social media, especially the ones with visual impairments. All of the misinformation interventions are designed for visually able users and studies with human participants have assumed that users can *access* these warnings without any need for assistance or reconfiguration. The experiences and exposures to misinformation amongst users who are low vision or blind have therefore been excluded in understanding the effectiveness, informativeness, usability, and deception-prevention of warning interventions, despite the fact that this group is equally active on social media as their visually able counterparts [54, 58, 103].

Past studies have shown that accessibility on social media is far from equitable when it comes to design consideration for people with visual disabilities [52, 100]. Given that the basic elements of social media participation is ridden with impediments, we suspected that the same will be the case with the misinformation warnings

as they have become an integral part of the overall experience on these platforms. As users who are low vision or blind do encounter misleading and false content in the same way as other visually able users, we identified an *exclusion* gap in aiding this group towards better truth discernment. To address this gap, we devised a study with 29 social media users who are low vision or blind and asked about their opinions, experiences, and recommendations for accessible misinformation warning interventions.

The three social media platforms of choice for this group of users were YouTube, Facebook, and TikTok as these are multimedia platforms that cater audible and textual content and not just graphical posts. Only a third of the participants were able to notice the misinformation warning labels on substantiated content on these platforms, despite using an assistive technology (e.g. screen reader, magnifier, large text, or color filters) to navigate the mobile interfaces. The remaining two thirds simply ignored the misinformation warning, despite the fact that 65% of all our participants have encounter these interventions across the social media ecosystem. Relative to the poor accessibility and omissions, only 31% of our sample said that that the misinformation warnings giving a bit of context. The lack of meaningful accessibility lead 72% of the participants to ignore the "nudge" towards a better truth discernment when accessing potentially misleading content on these platforms.

To report the findings from our study, we review the related work on the scientific "against misinformation" effort on social media in Section 2. Section 3 covers the implementation and reception of social media interventions against misinformation. Section 4 summarizes the difficulties and hardships low vision or blind users experience during their time on social media. Section 5 provides the methodological details of our study and Section 6 elaborates on the participants' experiences, opinions, and recommendations for *accessibility-inclusive* redesign of the misinformation warnings. We draw on our findings in Section 7 to discuss the implications for inclusive usability, truth, and future participation on social media when it comes to falsehoods and misleading statements. Finally, Section 8 concludes the paper.

2 AGAINST MISINFORMATION

The problem of "fake news" is not new. Misinformation was already in evidence two centuries ago, for example, in the political campaigning for the 1828 elections between Andrew Jackson and John Quincy Adams [17]. News and reporting were not impervious to the "fake news" temptations, and not long after, the New York Sun published the first 'Great Moon Hoax' of 1835 [66]. Fake news – be that propaganda, rumors, hoaxes, or any form of fabricated information – have a long historical precedence as information lacking truth and truthfulness saw multiplication both in the reach and the effect with the advent of global communication technologies such as the telegraph, radio, television, and the Internet [22, 56].

The relative lack of rules governing these communication technologies paved the way to rigging political elections, genocides, regime changes, health crises, and other disruptions through manipulative promulgation of misinformation [22]. While in the past the (fake) news consumption relative to these events was restricted to several outlets and agencies, social media recently created conditions where many news consumers feel entitled to choose or create their own "facts" [66]. The unprecedented threat of drowning factual and truthful information on social media united a multidisciplinary scientific front to work against the plight of fake news [8, 65, 76, 102, 110].

Epistemologists defined "fake news" as any information that exhibits both lack of truth and lack of truthfulness [43]. The lack of truth could result from falsehoods (e.g. the Pizzagate hoax [47]) or misleading statements derived from otherwise factual information (e.g. out-of-context COVID-19 death statistics [2]). The lack of truthfulness could result from propagation of falsehoods and/or misleading statements to either deceive (e.g. all of the Internet Research Agency's trolling tropes [18]) or with no regard to the truth (e.g. the Pope's endorsement of Donald Trump's presidential candidacy [82]). The absence of journalistic standards [50], lax moderation policies [102], and the natural human inclination for socialization [78] created circumstances in which fake news permeated on social media platforms beyond just opinion manipulation but also instigating polarization outside of the online realm.

2.1 Misinformation Concerns and Responses

People using these platforms started seeing content or "news" that aligns with their views and a growing variety of "facts" to choose from in forming their opinions and guiding their actions. Fake news now appeared as an information disorder worthy of attention of scientific community, prompting an immediate *against misinformation* action. As proving an intention to mislead (i.e. disinformation, computational propaganda [26]) or determining the degree of truth disregard (i.e. "bullshit" [25]) in reasonable time of practical relevance is a difficult endeavour, the fake news problem was largely refocused towards the production, propagation, and reception of the falsehoods and misleading statements immediate to social media users (i.e. misinformation).

The *against misinformation* effort, then, turned to uncover the misinformation producers, be that individual "fake news" mercenaries [82, 102] or state-sponsored factories, farms, or outfits [11, 18, 53]. The operational effect of misinformation was extended to encompass both public opinion manipulation but also abusing and antagonizing ordinary users [16]. As this extension was akin to *trolling* - the asocial behavior characteristic for the nascent hacker community in the 90's and early 2000's [79] - the producers of misinformation were collectively refereed to as "trolls" [36, 49].

Moving from the production to the propagation, the *against misinformation* effort tasked itself to hunt trolls on social media platforms and identify their tactics and tropes. Troll accounts were singled out for their ability to assume culturally relevant personas, establish homophilic relationships with unwitting users, and switch narratives at points conductive to user polarization, for example elections or referendums [24, 96]. So called "sock-puppet" accounts were also looked as the means of exploiting the algorithmic recommendation to amplifying misinformation narratives, orchestrated by "liking," "sharing," or "commenting" on selected false and misleading statements, offensive memes, and antagonistic hashtags between these accounts [4, 23]. The programmatic amplification of misinformation through "social bots" was also tracked over time and across platforms to anticipated interference attempts with the public opinion during destabilization-prone events like the COVID-19 pandemic [41, 42, 69].

Though offering actionable solutions to curb the propagation of misinformation on social media, the academic *against misinformation* effort was not adequately matched by mainstream platforms. Twitter, Facebook, Instagram, and other mainstream platforms like YouTube, implicated as enablers of fake news, did some account and content cleanup and started offering the option to "get the facts" about the contested topic in question [78]. Repeating the history itself, these seemingly open "social networks" – while enhancing the possibility for democratic discourse and promoting the ability of marginalized communities and individuals to engage in the public sphere – also allowed bad actors to spread misinformation and undermine both democracy and the rule of law [22].

2.2 Misinformation Effects

The risks of misinformation manifest with less than optimal truth discernment (i.e. the ability to correctly determine the difference between true and false truth-warranting claims) amongst social media users [64]. Early in the against misinformation effort, a common narrative emerged whereby politics drives susceptibility to fake news or that people are "better" at discerning truth from falsehood (despite greater overall belief) when evaluating politically concordant news [45, 95]. But this narrative didn't get traction as more and more evidence showed that the "better" truth discernment is associated with reflective reasoning, regardless of whether the news is consistent or inconsistent with their partisanship [6, 12].

With hard-to-avoid misinformation on social media, the *against misinformation* effort offered to help the users reflect about fake news. Cognitive interventions to "innoculate" against misinformation were developed as well as "debunking" strategies for dismissing falsehoods and misleading statements [51, 67]. Together with the availability of fact-checking services [98], these interventions were meant to nudge people people towards deliberate reasoning against misinformation (e.g. switching from System 1 to System 2 reasoning) [63] and overcome negative effects like emotional (rather than reflective) information processing [55].

3 MISINFORMATION INTERVENTIONS

Transferring the interventions *against misinformation* into practical implementations was predicated on the willingness of social media platforms to engage in content moderation in the first place, take input from (crowd-sourced) fact checkers, and deploy automated means of identifying candidate misinformation content [30, 59]. If alternative platforms like Gab, Gettr, or Rumble rejected any interventions in the name of free speech and preservation of social media as a marketplace of ideas [74, 108], mainstream platforms did change their policies to actively intervene with misinformation conduct [57, 93, 106]. One part of the platform intervention is *hard moderation*, where accounts spreading misinformation content are either suspended or altogether removed based on fact checks inputs and algorithmically determined omissions of truth and truthfulness [73]. Another part of the platform intervention is *soft moderation*, where warnings about potentially false or misleading content are substantiated underneath questionable posts, usually linked to credible or authoritative resources [75].

3.1 Hard Moderation

Moderation became the *de facto* misinformation intervention by social media platforms as they implemented governance mechanisms, albeit obtuse, to structure participation in a way to facilitate cooperation and prevent abuse [33]. One mechanism is the ability of the platform to monitor for misinformation misconduct and institute "bans" on accounts. Platforms banned and immediately removed bot, sock-puppet, and trolling accounts linked to farms and disinformation outfits [10], but also individual profiles that repeatedly voiced falsehoods or misleading statements (e.g. the accounts of Rep. Marjorie Taylor Greene or Robert F. Kennedy Jr. [2, 27]). Platforms also temporary suspended accounts that exhibited "inauthentic" behavior due to overly sharing misinformation content (e.g. pro-Bolsonaro influencers in Brazil [34]). Another mechanism is the flagging feature that allows users to report individual posts/accounts to the platform administrators or provide crowd-sourced context to misleading statements or falsehoods [94].

The hard moderation as a misinformation intervention received mixed reception among social media users [83, 99]. Described as a form of censorship or restraint on the users' voice, platforms have been accused of political bias [76], suppression of free speech [74], and intentional barring the visibility of users to the others on the platform (i.e. "shadowbanning") [71]. The so-called "deplatforming" or permanently banning of controversial public figures with large followings has also been the cause of discontent among some users, causing them to abandon mainstream platforms in favor of the alternative ones [3]. Interestingly, the hard moderation, in cases of accounts producing volumes of misinformation, decreased the need for subsequent soft moderation, as the interested for these narratives naturally subsided [68].

3.2 Soft Moderation

True to the effort for facilitating better truth discernment, platforms also implement mechanisms to "inform" or warn users about posts containing falsehoods or misleading statements. Substantiating posts with *misinformation warnings* comes in two flavors: (i) *covers* which obscure the misinformation and require users to click through to see the post [46]; and (ii) *labels* which appear under the post and do not interrupt the user or compel action [75]. Though both types of warnings insert a visual friction to draw the user's attention, their placement in the interface plays an important role in how they are utilized by users.

For some users, the placement of the misinformation warnings is irrelevant as neither of the visual frictions dissuade them from sharing and engaging with misinformation [62]. For other users, the covers work, but the labels don't, meaning that the interaction and visual friction of the covers is sufficiently potent for users to avoid a questionable post (but largely ignore the labels after been already exposed to falsehoods or misinformation statements) [72]. While the users' political self-identification had a limited effect in the truth discernment overall [65], interestingly, it briefly played a notable role in how users engage with misinformation warnings.

Studies show that Republicans or right-leaning users share misinformation labeled content much more than Democrats or left-leaning users [77, 107]. For a period, it was believed that the misinformation warnings frequently fail to reduce misperceptions among the targeted political ideological groups [15, 61, 72]. Also, in the

context of the COVID-19 pandemic, the labels resulted in a "belief echo," manifested as skepticism of adequate COVID-19 immunization particularly among Republicans and Independents [72]. However, further studies showed that this so-called "backfire effect" is not as serious as initially thought [48, 89, 101]. Evidence also emerged suggesting that social media users do heed misinformation corrections and platform warnings regardless of their political-self identification when a proper reflective context is present [75, 90].

Platforms, nonetheless, prefer the misinformation labels more. This is because the labels provide the advantages of mitigating their image of "overtly imposing", "biased," and "punitive" moderator amounts the users [44, 70], while maintaining high engagement by blending the labels with the user interface aesthetics (e.g. same colors, fonts, and obscure text) [75]. The misinformation warning labels, however, are asymmetrical in nature because the mere exposure to misinformation often generates a strong and automatic affective response, but the warning itself may not generate a response of an equal and opposite magnitude [29]. This is because the labels often lack meaning, have ambiguous wording, or ask users to find context themselves which is cognitively demanding and time consuming [20]. Evidence suggests that enhanced misinformation warning labels that provide either a context of the correction or are coupled with more salient visual frictions (e.g. an actual red flag superimposed as a watermark on a post) do work better for visually able social media users [75].

4 SOCIAL MEDIA USERS WHO ARE LOW VISION OR BLIND

4.1 Accessibility

Users who are low vision or blind participate on mainstream social media as much as the general population, with evidence suggesting that this user group receive more feedback (i.e. comments and likes) on average on the content they regularly post [103]. At least up until recently the engagement on mainstream social media, was mainly centered around sharing of images that, in turn, require rich and informative alt-text descriptions to provide full accessibility [58]. However, social media platforms are slow and reactive in addressing the essential needs for users who are low vision or blind, valuing accessibility less highly than the participation affordances specifically tailored for the majority of visually able users [54].

It took more than a decade for Facebook to make images accessible across all devices and it took repeated complaints for Twitter to fix their 280 character screen reader compatibility issues [100]. The audiovisual platforms like YouTube or TikTok fare no better amongst those users who are low vision or blind because there is no requirement or platform-provided alt-text descriptions, inline audio descriptions, or extended audio descriptions [52]. The visual impairments for users don't hinder the motivation to enjoy, create, and engage with YouTube or TikTok videos [60], but it requires a tedious trial-and-error search and non-trivial content preparation for upload, especially for short videos characteristic for TikTok that often contain embedded textual content which is difficult, or in some cases impossible to perceive with any assistive screen reading, magnification, or color contrast functionality [80].

To improve the accessibility, researchers have created various assistive tools such as web-browser extensions for automatic image alt-text captioning [31], GIF annotation systems [111], augmented video search functionality [52], and accessible video editing with audio-visual scripts [38]. But, even with the improved accessibility of captioning tools, users who are low vision or blind regularly encounter captions with missing details, hard-to-find videos, and difficulties in creating audio visual content. This group of users is not just excluded from a full participation on social media, but often has to deal with incongruent and incorrect captions to images and videos, effectively polluting their engagement with falsehoods and misleading statements from the get-go [54].

To the best of our knowledge, no universally adopted standard or guideline exists specifically for the purpose of designing inclusive misinformation interventions on social media. However, the Web Content Accessibility Guidelines (WCAG) are globally recognized and are the leading guidelines used in efforts of developing accessible experiences for all users of websites and applications in general [40]. According to these guidelines, and any application code should follow best practices and utilize HTML Semantics where possible [1] in order to enable intuitive text-to-speech translation. In cases where HTML Semantics cannot be used to achieve an inclusive experience, developers can utilize the World Wide Web Consortium's Web Accessibility Initiative Accessible Rich Internet Applications (WAI-ARIA) standard instead [39].

4.2 Experiences

In addition to the hindered engagement due to false captioning [54], low vision or blind users have a hard time to best capture the appearance characteristics that may convey the race, gender, and disabilities of those photographed, according to Bennett et al. [7]. Stangl et al. interviewed low vision and blind users to capture their preferences for best descriptions image captioning, finding that details such as emotions, image settings, as well as cultural and political contexts of the image are elements they want in alt-texts on social media in order to enjoy images meaningfully [87]. Similarly, Liu et al. found that larger portion of the content on YouTube, e.g. videos with only music and no audio description, metaphors, or referenced visual content in speech are inaccessible to them [52].

Testing a large video dataset for accessibility, Aydin et al. found the short videos, akin to the ones typical on TikTok, are hard to comprehend for users who are low vision or blind because the captioning provides little context about the events in the video [5]. Similarly, they found that many short videos contain music completely unrelated to the visual content, making it confusing or entirely inaccessible. Clutter due to disjointed social media feeds, non-negligible volume of adverts, and lack of topological interface organization, according to the empirical evidence analyzed by Whitney and Kolar, was found to be a significant factor that adds additional barriers in engagement on social media for users who are low vision or blind [100].

4.3 Misinformation

Encountering misinformation on social media for users who are low vision or blind hasn't received any scientific attention so far. As misinformative content was, and still is, floating on platforms with both false statements and misleading images and videos (e.g. political memes, doctored videos) [14, 19], it is likely that beyond

just encounters, users who are low vision or blind are equally engaged with it as their visually able counterparts. Users who are low vision or blind have pointed that they are equally exposed to subjectivity in information, unfamiliar data representation, and discrepancies between textual and image/audiovisual data on social media [84], making them the perfect target of rumors, conspiracy theories, and widely circulated falsehoods. Due to the incomplete accessibility of platforms, this group of users is particularly vulnerable to misconceptions, misinterpretation, and exclusion from identifying misinformation on social media.

It is unknown how users who are low vision or blind conceptualize misinformation in general and how they differentiate between inherently misinformation content and content that is misinformative due to misaligned accessibility (e.g. incorrect caption, missing details). It is also unknown to what this group of users benefits from hard moderation, how they conceptualize it, and whether they participate in actively flagging social media posts they believe constitute misinformation. One could presume that screen readers present both the covers and vocalize the misinformation labels, but there is no evidence that this delivery helps blind and visually impaired users towards a better truth discernment. In fact, the unintended effects of these misinformation interventions on social media on visually able users, is a large item of concern because neither the covers/labels were designed for accessibility friction nor their role in the overall platform experience is clearly communicated.

5 STUDY

We took it upon us to investigate the experiences and accessibility design recommendations of users who are low vision or blind relative to misinformation interventions on social media. We chose to focus on the platform warnings because they are specifically designed as *visual frictions*, which is a format likely to be a barrier towards a better truth discernment even with assistive technologies such as screen readers, magnifiers, or color filters. This design, perhaps unintentionally, fails to fully incorporate accessibility in the misinformation interventions, so we suspect that the opinions, needs, and design input from this group of users, representing 8% of the entire population in the US [86], have been largely omitted.

5.1 Research Questions

To explore this potential exclusion from misinformation interventions on social media, we invited 29 users who are low vision or blind to answer the following research questions:

- **RQ1:** What are the experiences of users who are low vision or blind with misinformation interventions on social media?
- **RQ2:** How do misinformation interventions on social media help users who are low vision or blind with truth discernment?
- **RQ3:** What design recommendations do users who are low vision or blind have for making misinformation interventions fully accessible and helpful towards better truth discernment?

It is important to note that we set no expectations regarding a potential difference between the experiences (RQ1), utilization (RQ2), and preferences (RQ3) of people who are visually able and people who are low vision or blind. Similarly, we set no expectation regarding any differences between the truth discernment levels of these two groups of social media users. There is no such evidence so far in the literature; Therefore, with our study we only challenged whether misinformation labels work for people who are low vision or blind (regardless of whether the labels are accessible or not and their demographic profiles, including political self-identification).

5.2 Sample

Our study was approved by the Institutional Review Board (IRB) of our university. We invited, through personal contacts, and snowballing sampling of users who are low vision or blind for a virtual interview session with open-ended questions (the interview script is listed in the Appendix). We sampled a population who were 18 years or older, from the United States, with significant vision loss per the definition for statutory blindness [86], and are regular social media users. Political positioning was not a selection criteria. We did however asked about their previous experiences with misinformation and all of our prospective participants reported they encountered various types of false content, conspiracies, rumors, etc. on social media. We used Zoom to conduct the interviews. Each interview was recorded, stored in a secure server, manually transcribed and then communicated with each interviewee to obtain an approval before we started the qualitative analysis.

We reached saturation and we concluded our recruitment at 29 participants in our sample. Each participant took roughly around 30 minutes for participation and was compensated with a $25 Amazon gift card. The demographics are given in Table 1 and the sample's visual profile is given in Table 2. The platform of choice for each participant is given in Table 3, indicating a strong preference for an audio-visual content on YouTube, Facebook, and TikTok. The participation in the study was not anonymous to us as researchers, but we anonymized the answers and removed any information that could potentially identify a participant. We allowed the participants to skip any question they were uncomfortable answering and we allowed for the participants the option to remove their answers if they wish too after the participation (none of them did).

Table 1: Demographic Distribution

Gender			
Female 16 (56%)	**Male** 12 (41%)	**Non-Binary** 1 (3%)	
Racial/Ethnic Self Identification			
White 14 (48%)	**Hispanic/latinx** 7 (24%)	**Asian** 5 (17%)	**Black** 3 (10%)
Political Self-identification			
Apolitical 5 (17%)	**Left-leaning** 12 (41%)	**Moderate** 7 (24%)	**Right-leaning** 5 (17%)
Age			
[18-29] 11 (38%)	**[30-39]** 8 (28%)	**[40-49]** 4 (14%)	**[50-59]** 2 (7%)	**[60+]** 4 (14%)
Education			
High-school 3 (10%)	**Undergraduate** 12 (41%)	**Graduate** 11 (38%)	**Doctorate** 3 (10%)

Table 2: Visual Disability Profile

Visual Self Identification			
Totally Blind	**Light Perception**	**Legally Blind**	**Low Vision**
4 (14%)	11 (38%)	11 (38%)	3 (10%)
Device			
iPhone	**Android**	**Windows PC**	
24 (83%)	1 (3%)	4 (14%)	
Assistive Technology			
Screen Reader	**Magnifier**	**Large Text**	**Color filters**
23 (79%)	4 (14%)	2 (7%)	2 (7%)

Table 3: Social Media Platform of Choice

Social Media Platform of Choice		
YouTube	**Facebook**	**TikTok**
14 (48%)	9 (31%)	6 (21%)

Our sample is diverse, balanced relative to gender, age, and political self-identification (screened at the end of the interview), and leaning towards the younger population. Past studies have used political self-identification to analyze the effects of truth discernment relative to politically concordant misinformation as well as the attitudes towards misinformation labeling. Participants in these studies were screened based on how they see themselves relative to the dominant political ideologies (left-leaning, moderate, right-leaning, apolitical) [76], partisan affiliation (e.g. Republican, Democrat, Independent) [74, 108], and their voting intention (e.g. for whom they plan to vote in the forthcoming elections) [35]. We chose to screen for participants' political positioning with a categorical variable assigned to the question "*When it comes to political positioning, you identify yourself as: left-leaning, moderate, right-leaning, apolitical?*, where *apolitical* indicated that the participant's political views were not aligned with the prevailing American political dichotomy.

5.3 Methods and Instrumentation

To capture the experiences with misinformation interventions on social media, we decided to use actual content that was labeled by each platform as potentially misleading and present it to the participants, shown in Figure 1. All posts were on the topic of COVID-19 and the updated origins of the virus, following the investigations conducted by the Department of Energy and the FBI [32]. We chose this topic as it was: (a) relevant at the period during which we conducted the study; (b) the participants were aware about the existence of the COVID-19 virus; and (c) COVID-19 is one of several explicitly noted topics in the policies of YouTube [105], Facebook [57], and TikTok [91] as targets for active moderation, including posts about the origins of the virus. We were, in the same time, aware that moderating COVID-19 content is a politically charged topic in the United States [92]. However, we selected the interventions shown in Figure 1 because content related to COVID-19 is the content our participants have or might encounter in future because platforms most actively moderate posts on this topic and out participants were also aware of the broader political context surrounding COVID-19 in general.

For YouTube, as depicted in Figure 1a, we used a video from the official Bloomberg news channel that was substantiated with a misinformation label titled "***COVID-19***" and urged the viewers to "*Get the latest information from the CDC about COVID-19*, offered a link to "*Learn more*," and added the option to "*See more resources on Google*" [9]. For Facebook, as depicted in Figure 1b, we used a video from the official Forbes news channel that was substantiated with a label titled "***Missing Context**: Independent fact-checkers say this information could mislead people*" and urged the users to "*See why*" [28]. For TikTok, as depicted in Figure 1c, we used a short video from the official Yahoo! news channel that was substantiated with a label reading "*Learn the facts about COVID-19*" [104]. We were not able to find content with covers on this topic on neither of the platforms even after an extensive search so we settled only on moderation with misinformation labels.

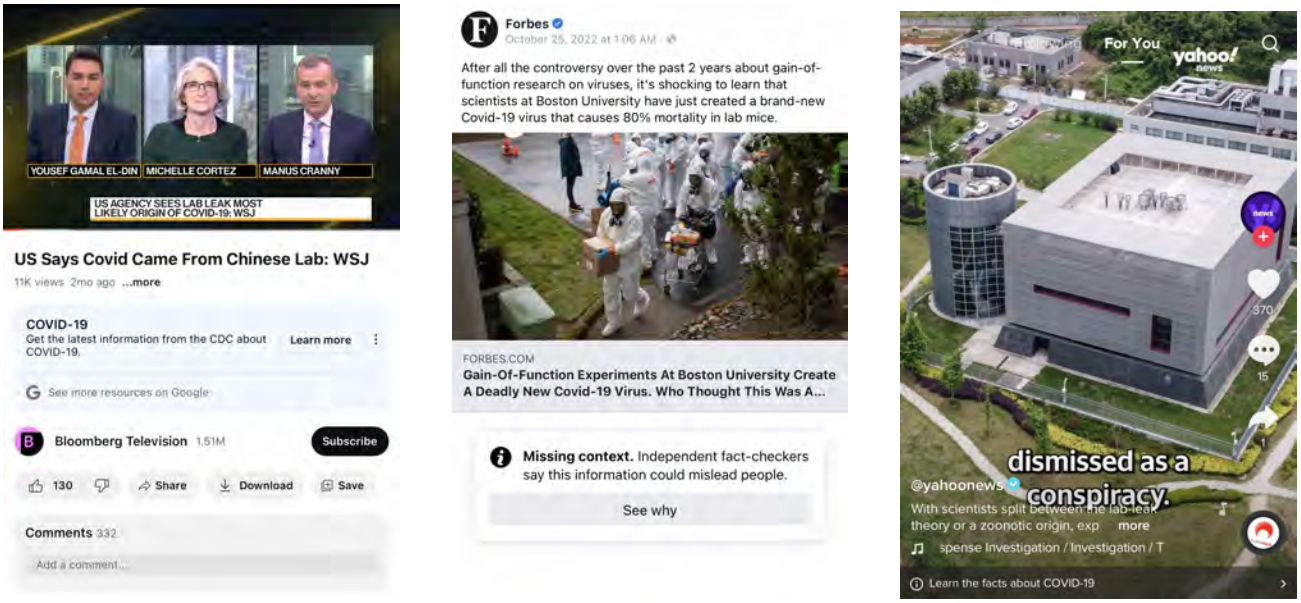
Each user was first asked about their platform of choice, then the corresponding link was shared with them over Zoom or an email if the participant expressed a need for better accessibility (each participant evaluated only one link). Participants were given unlimited time to access it with the assistive technology of their choice, and invited to speak about their experiences, opinions, and help with truth discernment from the misinformation warnings applied by these platforms. The qualitative responses were coded and categorized in respect: a) platform's accessibility (directly noticed or pointed out by the researchers if the participant did not notice the warning label); b) truth discernment (if the warning label is helpful in discerning the content's accuracy and reliability); c) opinions about the warning labels – truth discernment-wise and accessibility-wise; and e) accessibility design recommendations. The coding of the raw participants' responses was inductive and used the codebook provided in the Appendix with 23 codes in total.

Two independent researchers analyzed the approved interview transcriptions, achieving a strong level of inter-coder agreement (Cohen's $\kappa = .89$). We utilized a thematic analysis methodology to identify the themes and sub-themes most saliently emerging from the responses in our sample. The themes were summarized to describe the conceptualization, experiences, and opinions of misleading content in the view of the contemporary social media misinformation intervention. In reporting the results, we utilized verbatim quotation of participants' answers, emphasized in "*italics*" and with a reference to the participant as either **P#XZ** or [**P#XZ**], where **P** denotes **participant**, **#** denotes the **number** of the participant in the sample (ordered by the time of participation), **X** denotes their **political self-identification** identity (**L** - left-leaning, **M** - moderate, **R** - right-leaning; **A** - apolitical), and **Z** denotes their **platform** of choice in the study (**Y** - YouTube, **F** - Facebook, **T** - TikTok). For example, the **P15LY** reference links to **participant 15**, who politically self-identified as **left-leaning** and used **YouTube** as a platform of choice in the study.

6 RESULTS

6.1 Accessibility Experiences

Our first inquiry, in part, was regarding the experiences with the accessibility of the misinformation warnings on the social media platforms of choice selected by the participants in our study. Each participant was asked to open the associated content, shown in Figure 1, and use any of the assistive technologies they usually



(a) Youtube Post **(b) Facebook Post** **(c) TikTok Post**

Figure 1: Social Media Posts with Substantiated with Misinformation Labels

use to consume the content on their own pace. Then, we asked each of the participants whether they have noticed any platform-provided warnings, recommendations, or labels about the content they viewed as well as anything that stood out in the interface from their usual experience when accessing content. Eleven or 38% of our participants did notice the misinformation warnings directly on the platform, as shown in Table 4.

Table 4: Misinformation Label: *Direct* Access

Themes	Number of Responses
Located the label without difficulties	3
Located the label without difficulties and then expounded on their accessibility experience	4
Located the label without difficulties and then extrapolated their experience with past labels on social media platforms	3
Located the label with difficulties	1

Three of them did it without difficulties - two on YouTube and one on TikTok. Four (YouTube - 2, Facebook - 1, TikTok - 1) expounded on their accessibility experience, stating: "*So it took me a while to get down to it from an accessibility standpoint and in my own world, at this point, I probably would abandon the post because I don't want to click on a see why*" [**P10AF**]. Another three participants using each one of the platforms offered their past experiences with misinformation warning labels, stating: "*YouTube has like a little blurb, I don't know if it's like a Wikipedia thing or something, but tells you a little bit about the topic and I have noticed that with certain videos, not all of them.*" [**P7MF**]. Only the **P3RY** in this group had difficulties in navigating the interface, complaining that one "*cannot navigate into it with VoiceOver very easily, at least I have not found a way, so I am not a fan of these labels*".

Eighteen or 62% of our participants did not notice the warning labels, as shown in Table 5. Seven of them (YouTube - 6, TikTok - 1) were able to get back to and access it without difficulties, pointing out that the misinformation label has an obscured location "*under the title of the video*" [**P2AY**]. Other three participant from each platform expounded on their accessibility experiences, stating that it is confusing whether they "*are supposed to proceed with the label or the content first*" [**P20LT**]. The remaining eight participants did experience difficulties locating and accessing the misinformation labels (YouTube - 5, Facebook - 2, TikTok - 1), commenting that "*it is there, but it wasn't like so in my face that I like and I didn't have to interact with it in order to click play*" [**P25LY**].

We asked our participants to comment on the overall accessibility of the misinformation warnings, as shown in Table 6. Seven of them (24%) provided short affirmative answer that the content is accessible in a way that is read by the screen reader and seen when magnified/contrasted. Twenty or 69% of the participants felt that this "declarative" accessibility is without utility to them, making the misinformation labels *de facto* inaccessible. For example, participant **P11LF** stated that the "*little label is white on white down at the bottom of the post, so that would be very well easy to miss from a low vision perspective.*" The remaining two of participants (7%) agreed that the labels are practically inaccessible, commenting that "*the label is not the first think that VoiceOver is gonna read because you have to kinda scroll down a little bit to get to it*" [**P15LF**].

6.2 Truth Discernment

Our second inquiry was geared towards learning whether and how the misinformation warnings on social media platforms helped our participants with the truth discernment (i.e. if the labels made them consider the factual underpinning of the statements). As shown

Table 5: Misinformation Label: *Pointed-back* Access

Themes	Number of Responses
Returned to access it with their assistive technology without difficulties	7
Returned to access it with their assistive technology without difficulties and then expounded on their accessibility experience	3
Returned to access it with their assistive technology with difficulties	8

Table 6: Misinformation Label: *Accessibility* Opinion

Themes	Number of Responses
Accessible, default (screen read, magnified, contrasted)	7
Accessible, but without utility (ignored, intelligible); practically ***not*** accessible	20
Not accessible	2

in Table 7, only eight or 28% heeded the misinformation warning (YouTube - 2, Facebook - 4, TikTok - 2) and stated that "*it made [them] second guess*" [**P6**] and "*helped them as a cue not to watch similar videos*" [**P15**]. Participant **P27LY** reasoned:

> *If I'm taking a step back and evaluating this, it seems to me like the platform is trying to take an objective stance and provide guidance that has been provided by the government rather than swing one direction or the other, really being neutral in regards to the subject matter, if neutrality is sharing information that the CDC is providing. So when I see something like this from a platform trying to balance out the issue as it relates to the content it kind of makes me question the content, more.*

Table 7: Misinformation Label: Truth Discernment

Themes	Number of Responses
The label did help the participant with truth discernment	8
The label ***did not*** help the participant with truth discernment	14
The label ***did not*** help and the participant opined on the truth discernment effect of misinformation warnings	7

Fourteen or 48% of the participants (YouTube - 7, Facebook - 4, TikTok - 3) stated that the misinformation labels did not help them with the truth discernment. Speaking about YouTube's label, participant **P8RY** simply said: "*It didn't helped me because it really did not give me any information itself besides saying - here are couple of links, if you want to go to them*". Speaking about Facebook's label, participant **P19LF** said that it did not change anything for them because the label "*says independent fact checkers but it doesn't say which independent fact checkers.*" Speaking about TikTok's label, participant **P20LT** felt that "*the label itself doesn't make the content it substantiates more or less reliable for me, it just means there's more about it somewhere else*". The remaining seven (24%) of the participants that did not find the misinformation labels useful offered their opinions on the truth discernment effect in general. Participant **P3RY** commented:

> *In my experience, from what I've seen, these labels are bots. Really, it's like they Google has incorporated this kind of automatic technology where it could flag anything it wants, and just like that, put the label on it. As a user, honestly, I'll thumb my nose at it because it kind of upsets me that they're trying to control people's opinions, and I'm not one to enjoy that.*

We also asked the participants to provide their general opinion on labels interventions against misinformation on social media, as shown in Table 8. Four of them (YouTube - 2, Facebook - 2) thought that the misinformation labels are in the way of accessing and consuming actual content, stating that "*they don't need periodically something to say this has been proven true and this is why, and this has been proven false and this is why*" [**P28LF**]. Another four (YouTube - 2, Facebook - 2) believed that the labels are intrusive and promote the platform "hidden" narrative. Participant **P29RY** stated:

> *I'm totally against these labels as they take away my right to free speech. I have a constitutional right to say what I believe, whether it's against what people like to hear or not. I'm willing to listen to their stupid points of view and they should listen to my point of view. That is freedom, that is free speech and anytime they start putting up hindrances to that makes me very, very irritated, like very angry. These government and private companies have no business interfering.*

Table 8: Misinformation Label: *General* Opinion

Themes	Number of Responses
Participant opines that the labels are in a way of accessing content	4
Participant opines the labels are intrusive and promote the platform "hidden" narrative	4
Participant opines that the labels are wrongly substantiated	2
Participant opines that the labels are indistinguishable and therefor don't work	10
Participant opines the labels give a bit of context about the content they are substantiated to	9

Two of the participants were on the opinion that the misinformation labels are wrongly substantiated on both YouTube and TikTok videos, saying that "*some obvious stuff that is not a fact just random people talking gets a label for no reason*" [**P6LT**]. Ten of the participants pointed out that by being inaccessible, the misinformation labels don't really make any difference for both visually able and impaired or blind users. Participant **P20RT** was on the opinion that "*usually platforms don't put anything on there that's going to make you think that what you're looking is not like 100% reliable, otherwise it would make the person question why they were using that platform.*" The remaining nine participants expressed positive opinions about the warnings, stating that "*the labels have the potential to immediately put some people on red alert when the post they are on is not fully accurate*" [**P12AF**].

6.3 Accessibility Redesign Recommendations

Our third inquiry collected the participants' recommendation about how would they redesign the misinformation warnings in order to be meaningfully accessible, and with that, achieve full *inclusivity* for all social media users. Here, we aimed to collect participants' recommendations for both improvement of accessibility and truth discernment on social media. As shown in Table 9, six of our participants or 21% (YouTube - 4, Facebook - 1 , TikTok - 1) of the participants wanted the warning labels entirely removed from the platforms. For example, participant **P2AY** simply stated: "*I am not interested in warnings nor in looking up the information they provide so I don't want them when I access content on YouTube.*"

Table 9: *Accessibility* Redesign Recommendations

Themes	Number of Responses
Participant wants the labels entirely removed from the platform	6
Participant prefers a ***verbose cover***	11
Participant prefers a ***verbose before-content label***	13
Participant prefers a ***before-content audio signal or vibration in addition to a cover or a label***	4
Participant prefers a ***stark contrast, bold/large font, standout colors different from the platform's aesthetics***	4

Eleven or 38% of our participants demanded a **verbose cover** instead of a misinformation warning label, providing a substantive amount of text sufficient to capture the purpose of the label, the relevance regarding the content, and allow for the screen readers to read it in a meaningful way for blind and visually impaired users. Participant **P3RY**, suggesting a redesign of the YouTube labels, started: "*I suggest it is some kind of a cover or a pop up before the video if you want to get more details; Or if you want to continue, I think that could also work that you need to actually press play to do so*". Participant **P17MY** added that "*the main thing is really just knowing that it's there and I guess the best way would be to have a dialogue that appears that's more intrusive and users have to click or say something to get to the video.*" Participant **P25LY** justified these design recommendations from a practical point of view, suggesting "*a pop-up box because when you use a screen reader, blind users often speed up the speech so fast so we don't always take in what we read – similar to sighted people scrolling – so let's do something obviously different about the label and make it more verbose to stand out.*"

The pop-up idea was also recommended by the users on Facebook, suggesting an "*automatic reading when you access a post*" in order to ensure that users won't miss it, having the option for users to say "*okay to continue*" [**P5LF**]. Participant **P19LF** provided a useful analogy as a design recommendation for Facebook labels: "*I think it should come up as an notification that you press when you go to the video – just like how on my phone says low battery – and then I have to press the close button for it to go away, then you would know for sure it was there.*" Similarly for TikTok, participant **P22RT** thought "*it would be more accessible if the label is larger, maybe at the top of the screen, with a bigger font or something, something to bring more attention to it, perhaps playing the whole time the video is playing.* Participant **P6LT** added:

> *I feel like it is so easy to just scroll to the next video without recognizing the warning or having time to learn the facts. So maybe if you scroll, and then, a screen pops up, and it will feel like something people have seen before, for example ads. Maybe it should ask the person if this post was interesting to them and then the user can say yes or skip (I usually just say, skip). But then I feel like that would give you like one extra time to like the post too. Also, I am worried that if there were a lot of warnings on a lot of different posts that would get kind of annoying. But that's maybe be just a way to do it only for COVID-19. Or maybe you could turn the labels like a feature in the settings or something, although I don't know if I would actually want that.*

Along these lines, thirteen or 45% of the participants preferred the misinformation label, only to be **before** the content instead of after, as it is substantiated now on all three platforms. Several participants wanted "*standardized words or headings that explicitly read 'content notice' or 'label' so they at least can get a pattern for something that has a point to pay attention to and it's not just an ad*" [**P25LY**]. Participant **P23LY**, to this point, gave instructive design recommendations based on what does not work with the current implementation of the labels on YouTube:

> *The label is not designated as an interactive element, meaning it does not say link, it doesn't say button, it doesn't say drop-down, it doesn't say combo box. I don't know how to interact with this element or what's going to happen. As I continue to swipe right, it just says COVID-19, I think that's an interactive element. I think that if I click that something would happen but I don't get that from the screen reader output. And then it says learn more. Then when I continue to swipe right it says action menu. Action menu for what? As it comes down to it, the labels should be designated as separate interactive elements with a functionality attached to them, for example a combo box with a link that gives facts about the content. This will avoid confusion because now I am not sure exactly what's going to happen when I interact with the area specific to the label on YouTube.*

For Facebook, participants proposed the labels to be "*formatted more like a status notification, outside of the content*" [**P7MF**] so to stand out more as soon as the page loads instead of being embedded in the page. For low vision users, participant **P11LF** suggested "*something that really jumps out at you, like black on yellow contrast or at least big red border around it, because now it blends with the background on Facebook for someone with a low vision.*" For blind users, participant **P12AF** invoked the analogy with system notifications: *So if I were voice over it, it should automatically read that to me first and then offer a voice over selection and cursor so I can go to the video.* For TikTok, participants also invoked the lack of designated interactive elements dedicated to labels, with **P18MT** noting:

> *So all of the buttons would need to be clearly designated. I might implement headings or other aspects. If there was like a heading, for example, then I can navigate the page by a heading, and I could find the content label if it was a heading. So just kind of implementing*

> *those like easy navigate, making sure buttons are designated, making sure links are designated, having headings implementing those kind of aspects to make screen reader navigation more simplified, and then whatever issue is going on where, when you're scrolling for the screen, reader, the video also scrolls that would need to be worked on. I also think it would be super useful. If when you start the video, your focus in the screen just automatically goes to that content label so that you hear voice over reading it.*

Four of the participants extended the current methods of implementing misinformation warnings with a suggestion to add an **audio signal** or **vibration** in addition to a label that is substantiated **before** a post, as a way of achieving full accessibility. This idea of *audio/vibration friction* is analogous to the *visual friction* of the warnings, and in the opinion of participant **P12AF**, this should be a "*distinctive sound combined with vibration from the others the assistive technology makes so it alerts*" the user that it is something that should be paid attention to. The remaining four participants wanted a **visually salient** notifications instead of an indistinguishable text. For example, participant **P13MT** noticed that the label on TikTok "*doesn't have an underline formatting like a link but an arrow, a practically impossible situation to know that there is more behind that obscure text,*" and suggest a putting the label on the vertical right side with start contrast where the engagement elements reside, instead of the bottom of the screen.

7 DISCUSSION

For such an important aspect of (mainstream) social media participation, the current approach for soft moderation of misinformation largely fails to account for the needs of users who are low vision or blind. In the absence of platform-provided rich alt-text, inline/extended audio descriptions and captions, it is hard to imagine how users can basically engage with a content, let alone consider any "assistive" elements in the interface intended to help them with determining the factual underpinning of the same content. Several of our participants noted that none of the platforms have dedicated designation for the misinformation labels as separate interaction elements that the screen readers can convey back to them when accessing a particular video. This is a serious omission that, even in the presence of assistive captioning, search, or video-scripting tools [38, 52], ultimately excludes users who are low vision or blind from the benefit of fact-checked information and truthfulness context.

This omission exacerbates the chronically poor inclusivity of users who are low vision or blind on social media as it creates a *double misinformation disadvantage* – incongruent captions could misinform a user [54] but an inaccessible misinformation label could well do it too in the same time. The equal opportunity for truth discernment, as our results show, is practically nonexistent on YouTube, Facebook, and TikTok as almost three quarters of our sample did not act in response to the interaction "friction" of the misinformation labels. Restrictive as such, the labels reinforce the exclusion practice of social media platforms of inattention to the preferences of users who are low vision or blind for conveying detailed as well as cultural and political contexts of the content they host [87].

We specifically selected videos where the content is only verbalized statements superimposed on known images in order to address the inaccessibility to videos with only music and no audio description, metaphors, or referenced visual content [5, 52]. Even with this "advantage," the misinformation warnings failed to complement the statements with contextual warning about the relative truth and truthfulness, even if they label text was captured and read back to the participants or noticed through magnification/contrast. The unavoidable clutter of disjointed social media feeds, non-negligible amount of adverts, and lack of topological interface organization [100], clearly added to the indistinguishability of the misinformation labels, forcing us to explicitly point 62% of our participants back to them so they find them on the platform pages.

As in the past studies with visually able users [62, 72], we also found evidence that the labels don't work for their low vision or blind counterparts. We did ask our participants about the accuracy and reliability of the statements in the videos, but many of them refrained from explicitly agreeing or disagreeing with the new evidence about the origin of the COVID-19 virus. They did, however, explicitly state if the misinformation warning helped them or not change their prior stance or form a new one after being exposed to the videos. Therefore, we weren't able to actually measure for any effects of backfiring [48, 89, 101] or belief echoing [72]. However, we identified evidence of exclusion by the platforms' decision to implement labels that lack meaning, are ambiguously worded and ask users to find context themselves outside of the platform [20].

It is important to note that we noticed a tension between the participants' preferences for accessibility redesign recommendations of the misinformation warning labels. While many of them provided concrete and actionable suggestions for making these warnings more prominent and harder to ignore, there were participants that wanted these warnings entirely removed from the platform. We didn't investigated this tension more in-depth in our interviews, but it is a worthy aspect to pay attention to, particularly in learning in what way the participants contrast the accessibility of the labels with their perceived importance of the misinformation warning labels for truth discernment in the first place. For example, most of the participants in our study saw the problem with the misinformation warnings as part of a general poor experience with web and application design but there were few participants that saw the "interventions by the platform" as a deeper issue, one related to the right of free speech [74]. Many of the participants who were supportive of better misinformation warning labels acknowledged that their suggestions might also work well for visually able users and therefore are worthy of implementing them as a general improvement not just for screen readers but for a better overall UI experience for all users on these platforms.

In this context, the proposed accessibility redesign recommendations might work well for better truth discernment for all social media users as past evidence suggest that the "better" truth discernment is associated with reflective reasoning [6, 12]. To induce the reflection, or move from System 1 to System 2 reasoning, for example, social media platforms might allow users to select the type of warnings they want to see in their feeds, allowing them to pick a verbose covers, verbose before-content labels, audio signals and vibrations in addition labels, as well as different color templates for the warnings than the default platform colors. (the

user customization is important as many users might find some of these interventions annoying or prohibitive of a truly immersive experience on the platforms). Equally, the accessibility enhancements proposed in our study could be useful for inoculation against misinformation as the verbosity might be used to actually help both prebunk and debunk some obvious falsehoods [51]. Past evidence with visually able users suggests that a verbose misinformation label is indeed a helpful intervention to stop and think better about a false or misleading content on social media [75]. The verbosity, together with an audio or vibration, could serve as an alternative friction to the default visual one that assist some users – both visually able and low vision or blind – to overcome negative effects like emotional (rather than reflective) information processing [55].

Whether this accessibility recommendations will in fact result in better truth discernment or misinformation avoidance for both groups of users we can't say at this point. What we could say is that our study not just uncovers a problem of accessibility exclusion in the way warning labels are currently implemented as visual frictions, but also that the labels themselves might be indistinguishable and hardly of use for visually able users too. We plan to dedicate our future work in exploring whether there are any tensions or interaction/accessibility conflicts between the needs of these two groups, how these needs actually affect the truth discernment performance between these groups, as well as how enriched frictions as proposed in the current study could shape the future experience on multimedia rich social media platforms. Our study also raised the issue of the general need for better accessibility when it comes to security warnings in general. For example, email or browser warnings about malicious or phishing websites are also designed mainly as visual frictions with a color contrast but to what extend this is useful for users who are low vision or blind to avoid being phished is largely unknown. Therefore, we are committed in broadening our inquiry for inclusive security to understand the accessibility and interaction needs of this group for a larger set of security indicators, notifications, and frictions.

Our findings have clear implications for screen reader providers relative to how they should be improved in the context of inclusive security, both regarding misinformation warnings as well as general security warnings. It is unlikely that YouTube, Facebook, or TikTok will change their policies and require platform-provided alt-text descriptions, inline audio descriptions, or extended audio descriptions [52, 54]. Therefore, the onus is on screen readers to work on improvements in the way they capture the warning label text, enrich it with alt-text or extended audio descriptions. None of the screen readers we encountered in our study provided an alt-text about the warning icon that substantiated the warning labels on Facebook and TikTok. Equally, none of the screen readers captured the difference of the text font between the YouTube general textual content (12 points) and the textual content in the misinformation warning label (title: 11 points, text: 10 points). Also, the screen readers simply synthesized the text in an incongruent and incorrect caption to the videos as they weren't able to notify the users that each warning also contained a link that they might open if they want to "see why," or "learn more" about the facts.

Our findings raise the issue of non-compliance with the Web Content Accessibility Guidelines and the absence of best practice adoption of HTML Semantics as well as WAI-ARIA standards to provide guideline specifically for the purpose of designing inclusive misinformation interventions on social media. For example, many of our participants mentioned that the labels should be made as designated interactive elements so the screen readers can announce them separately from the other elements on the page. This will also help with the ability for low vision or blind users to easily follow up with the web link embedded in the warning that otherwise is not accessible in the current label formatting.

7.1 Ethical Consideration

Ethical concerns do arise when dealing with misinformation, as the probability of harmful implication is non-negligible within a pluralistic social media population including the community of users who are blind and visually impaired. Exposing participants to misleading statements, even in controlled settings and with an extensive debriefing, runs the risk of conceiving or perpetuating any misconceptions about the origins of the COVID-19 virus as a developing and actively debated topic on social media. Soft moderation, at least in our view, is a form of honest communicative action rather than an authoritative and absolute determination of truth, and as such, beneficial to as an intervention to misinformation.

Not all participants shared this view and saw us as a direct threat to their right of free speech, operating on the behalf of YouTube, Facebook, and TikTok. We employed lengthy explanations to ensure they accept our political impartiality, independence from any platform, and our ultimate goal of making misinformation warnings inclusive and accessible to a population that is regularly ignored and excluded from design deliberations and interface implementations. We were careful not to appear as staunch advocates of soft moderation, as several of our users were adamant that misinformation warnings should go away altogether from social media. Therefore, we position ourselves as honest listeners and facilitators of a democratic participation in an accessibility-inclusive design endeavour that allows for remediation of concerns in such instances.

7.2 Limitations

Our research project comes with several limitations. The scope of the study is limited to English-speaking social media users in the United States and only on three platforms. While the choice of the platforms is justifiable – given the natural fitting of a multimedia content compared to a graphic-heavy one – our participants could well provide different opinions, recommendations, and perhaps get to a better discernment if other platforms were considered, for example Twitter. Twitter's birdwatch program for crowdsourced misinformation labeling [94], albeit text heavy, does provide verbose labels that could be sufficiently long enough to achieve the necessary friction intended.

Next, we were limited to one topic of presumably misleading content related to COVID-19. Other topics, for example election rumors, conspiracy theories, or downright falsehoods could well be received differently by the participants with the current format of misinformation interventions being sufficient for a consensus against such a content. Here, we would like to add a note of caution for interpreting our results in the context of the participants self-reported political positioning to avoid speculation. We only reported the political positioning to capture more of a rich insight

into the answers to the third research question, provided that this demographic variable has been shown to play role in how users engage with labels on social media in the past. By no means we intent to stereotype or judge these users concerning their self-reported political stances as it might very well be that they didn't factor into how our participants saw the task of accessibility redesign. We were also limited to only misinformation labels, but not covers, as we were unable to find any such warnings on neither of the three platforms in reasonable time for completing the study.

The current state of the content moderation policies on YouTube, Facebook and TikTok is something we assumed as an operational reference for this study. Any future changes that do not deem the videos we chosen as misinformation candidates would entail a reassessment of the inquires relative to the labeling rules. In as much as we tried to limit any confounding factors when selecting the videos, there certainly are elements that factor in the truth discernment, if not the accessibility, among the participants such as the source of the videos, length, personas appearing, the surrounding recommendations, and even the comments/engagement sections attached to the videos. A limitation also comes from the sampling method we chose to our convenience, as other users who are low vision or blind and other samples might provide results that differ from the once we obtained as there is little insight into general sampling and sample-related differences when users are broadly queried about misinformation on social media.

The current accessibility support on YouTube, Facebook, and TikTok and the format of the misinformation labels is also another limitation. A change in the support, the format, the placement, and the design will of course require reconsideration our results to ensure the demands for inclusion are actually met by these platforms. Equally, a limitation comes from the current assistive technologies that were built in the operating systems our participants use. Any future developments of new assistive technologies for users who are low vision or blind might change how the users experience misinformative content and the substantiated interventions, and with that, affect the overall findings. Lastly, we left our participants sufficient time and support to meaningfully engage with the videos through the assistive technology of their choice, but that could have, nonetheless, been insufficient for our participants to formulate a more informed opinion or recommendation.

7.3 Future Work

The findings of our study are sufficiently informative for future action on several fronts. First, we plan to approach and work with the platforms in the study and any other interested to make marked improvements of the current soft moderation approaches towards an acceptable accessibility with an immediate effect. Second, we plan to broaden our research work to other divisive topics that are soft moderated and also test the accessibility and truth discernment effectiveness of misinformation covers. Third, we plan to use the accessibility-inclusive design recommendations from our participants to develop assistive tools for misinformation interventions specifically designated to participants who are blind and visually impaired. Testing and continuously improving these assistive tools not just on social media but elsewhere online, e.g. chat-rooms like Discord, is also on our research agenda and we sincerely welcome any collaboration offers, suggestions, and input that will ultimately benefit the community of low vision or blind users.

8 CONCLUSION

Misinformation undoubtedly shapes the way all people – visually able, blind, or with visual impairments – make decisions in their daily lives. As misinformation regularly floats on social media, aiding these decisions with interventions about possible falsehoods and misleading statements became a necessity to maintain a meaningful level of user engagement. Our findings indicate that these interventions are hardly accessible for users who are low vision or blind, precluding an equal opportunity for participation. Marginalized as such, our participants proposed valuable improvements for inclusive and accessible misinformation labels, which we sincerely hope will materialize without a delay on all social media platforms.

REFERENCES

[1] [n.d.]. Semantics - MDN Web Docs Glossary: Definitions of Web-related terms | MDN. https://developer.mozilla.org/en-US/docs/Glossary/Semantics

[2] Davie Alba. 2022. Twitter Permanently Suspends Marjorie Taylor Greene's Account. https://www.nytimes.com/2022/01/02/technology/marjorie-taylor-greene-twitter.html

[3] Shiza Ali, Mohammad Hammas Saeed, Esraa Aldreabi, Jeremy Blackburn, Emiliano De Cristofaro, Savvas Zannettou, and Gianluca Stringhini. 2021. Understanding the effect of deplatforming on social networks. In *13th acm web science conference 2021*. 187–195.

[4] Dennis Assenmacher, Lena Clever, Lena Frischlich, Thorsten Quandt, Heike Trautmann, and Christian Grimme. 2020. Demystifying Social Bots: On the Intelligence of Automated Social Media Actors. *Social Media + Society* 6, 3 (2020), 2056305120939264. https://doi.org/10.1177/2056305120939264 arXiv:https://doi.org/10.1177/2056305120939264

[5] Ali Selman Aydin, Yu-Jung Ko, Utku Uckun, IV Ramakrishnan, and Vikas Ashok. 2021. Non-Visual Accessibility Assessment of Videos. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 58–67.

[6] Bence Bago, David G Rand, and Gordon Pennycook. 2020. Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of experimental psychology: general* (2020).

[7] Cynthia L Bennett, Jane E, Martez E Mott, Edward Cutrell, and Meredith Ringel Morris. 2018. How teens with visual impairments take, edit, and share photos on social media. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12.

[8] Sven Bernecker, Amy K Flowerree, and Thomas Grundmann. 2021. *The epistemology of fake news*. Oxford University Press.

[9] Bloomberg Television. 2023. US Says Covid Came From Chinese Lab: WSJ. https://www.youtube.com/watch?v=Z8J7zWMF__8

[10] Brandon C Boatwright, Darren L Linvill, and Patrick L Warren. 2018. Troll factories: The Internet Research Agency and state-sponsored agenda building. *Resource Centre on Media Freedom in Europe* 29 (2018).

[11] David A. Broniatowski, Amelia M. Jamison, SiHua Qi, Lulwah AlKulaib, Tao Chen, Adrian Benton, Sandra C. Quinn, and Mark Dredze. 2018. Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. *American Journal of Public Health* 108, 10 (2018), 1378–1384. https://doi.org/10.2105/AJPH.2018.304567

[12] Michael V. Bronstein, Gordon Pennycook, Adam Bear, David G. Rand, and Tyrone D. Cannon. 2019. Belief in Fake News is Associated with Delusionality, Dogmatism, Religious Fundamentalism, and Reduced Analytic Thinking. *Journal of Applied Research in Memory and Cognition* 8, 1 (2019), 108–117.

[13] Lewis Carroll and Martin Gardner. 1993. Jabberwocky. In *The Annotated Alice: Alice's Adventures in Wonderland & Through the Looking Glass* (reprint edition ed.). Random House Value Publishing, New York : Avenel, N.J, 191 – 197.

[14] Matthew Childs, Cody Buntain, Milo Z. Trujillo, and Benjamin D. Horne. 2022. Characterizing Youtube and Bitchute content and mobilizers during us election fraud discussions on twitter. In *14th ACM Web Science Conference 2022*. 250–259.

[15] Katherine Clayton, Spencer Blair, Jonathan A Busam, Samuel Forstner, John Glance, Guy Green, Anna Kawata, Akhila Kovvuri, Jonathan Martin, Evan Morgan, et al. 2019. Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. *Political Behavior* (2019), 1–23.

[16] S Coleman. 2018. The elusiveness of political truth: From the conceit of objectivity to intersubjective judgement. *European Journal of Communication* 33, 2 (2018), 157–171. https://doi.org/10.1177/0267323118760319

[17] James W Cortada and William Aspray. 2019. *Fake news nation: the long history of lies and misinterpretations in America.* Rowman & Littlefield.

[18] Renee DiResta, Kris Shaffer, Becky Ruppel, David Sullivan, Robert Matney, Ryan Fox, Jonathan Albright, and Ben Johnson. 2019. The tactics & tropes of the Internet Research Agency. (2019).

[19] Marc J. Dupuis and Andrew Williams. 2019. The Spread of Disinformation on the Web: An Examination of Memes on Social Networking. In *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI).* 1412–1418. https://doi.org/10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00256

[20] Ullrich KH Ecker, Stephan Lewandowsky, and David TW Tang. 2010. Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & cognition* 38, 8 (2010), 1087–1100.

[21] Ullrich K. H. Ecker, Stephan Lewandowsky, John Cook, Philipp Schmid, Lisa K. Fazio, Nadia Brashier, Panayiota Kendeou, Emily K. Vraga, and Michelle A. Amazeen. 2022. The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology* 1, 1 (2022), 13–29.

[22] Harold Feld. 2020. From the telegraph to Twitter: The case for the digital platform act. *Computer Law & Security Review* 36 (2020), 105378.

[23] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. The Rise of Social Bots. *Commun. ACM* 59, 7 (June 2016), 96–104. https://doi.org/10.1145/2818717

[24] Paolo Fornacciari, Monica Mordonini, Agostino Poggi, Laura Sani, and Michele Tomaiuolo. 2018. A holistic system for troll detection on Twitter. *Computers in Human Behavior* 89 (2018), 258 – 268. https://doi.org/10.1016/j.chb.2018.08.008

[25] Harry G Frankfurt. 2005. *On bullshit.* Princeton University Press.

[26] Deen Freelon and Chris Wells. 2020. Disinformation as Political Communication. *Political Communication* 37, 2 (03 2020), 145–156.

[27] Sheera Frenkel, Davey Alba, and Raymond Zhong. 2020. Surge of Virus Misinformation Stumps Facebook and Twitter. https://www.nytimes.com/2020/03/08/technology/coronavirus-misinformation-social-media.html

[28] Frobes. 2023. Gain-Of-Function Experiments At Boston University Create A Deadly New Covid-19 Virus. Who Thought This Was A Good Idea? https://www.facebook.com/forbes/posts/after-all-the-controversy-over-the-past-2-years-about-gain-of-function-research-/10160609782322509/

[29] Bertram Gawronski. 2021. Partisan bias in the identification of fake news. *Trends in Cognitive Sciences* 25, 9 (2021).

[30] Tarleton Gillespie. 2018. *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media.* Yale University Press.

[31] Cole Gleason, Amy Pavel, Emma McCamey, Christina Low, Patrick Carrington, Kris M Kitani, and Jeffrey P Bigham. 2020. Twitter A11y: A browser extension to make Twitter images accessible. In *Proceedings of the 2020 chi conference on human factors in computing systems.* 1–12.

[32] Michael Gordon and Warren Strobel. 2023. Lab Leak Most Likely Origin of Covid-19 Pandemic, Energy Department Now Says. https://www.wsj.com/articles/covid-origin-china-lab-leak-807b7b0a

[33] James Grimmelmann. 2015. The virtues of moderation. *Yale JL & Tech.* 17 (2015), 42.

[34] Anatoliy Gruzd and Philip Mai. 2020. Going viral: How a single tweet spawned a COVID-19 conspiracy theory on Twitter. *Big Data & Society* 7, 2 (2020), 2053951720938405.

[35] Andrew M Guess, Michael Lerner, Benjamin Lyons, Jacob M Montgomery, Brendan Nyhan, Jason Reifler, and Neelanjan Sircar. 2020. A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences* 117, 27 (2020), 15536–15545.

[36] Claire Hardaker. 2015. "I refuse to respond to this obvious troll": An overview of responses to (perceived) trolling. *Corpora* 10, 2 (2015), 201–229. https://doi.org/10.3366/cor.2015.0074

[37] Jonathan Hardy. 2021. Media systems and misinformation. In *The Routledge Companion to Media Disinformation and Populism.* Routledge, 59–70.

[38] Mina Huh, Saelyne Yang, Yi-Hao Peng, Xiang'Anthony' Chen, Young-Ho Kim, and Amy Pavel. 2023. AVscript: Accessible Video Editing with Audio-Visual Scripts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.* 1–17.

[39] W3C Web Accessibility Initiative (WAI). [n.d.]. WAI-ARIA Overview. https://www.w3.org/WAI/standards-guidelines/aria/

[40] W3C Web Accessibility Initiative (WAI). [n.d.]. WCAG 2 Overview. https://www.w3.org/WAI/standards-guidelines/wcag/

[41] Peter Jachim, Filipo Sharevski, and Emma Pieroni. 2021. TrollHunter2020: Real-time Detection of Trolling Narratives on Twitter During the 2020 US Elections. In *International Workshop on Security and Privacy Analytics 2021* (Online, USA) *(IWSPA '21).* Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3445970.3451158 https://doi.org/10.1145/3445970.3451158.

[42] Peter Jachim, Filipo Sharevski, and Paige Treebridge. 2020. TrollHunter [Evader]: Automated Detection [Evasion] of Twitter Trolls During the COVID-19 Pandemic. In *New Security Paradigms Workshop 2020* (Online, USA) *(NSPW '20).* Association for Computing Machinery, New York, NY, USA, 59–75. https://doi.org/10.1145/3442167.3442169

[43] Romy Jaster and David Lanius. 2021. Speaking of Fake News. *The epistemology of fake news* 19 (2021).

[44] Shagun Jhaver, Darren Scott Appling, Eric Gilbert, and Amy Bruckman. 2019. "Did You Suspect the Post Would Be Removed?": Understanding User Reactions to Content Removals on Reddit. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 192 (Nov. 2019), 33 pages. https://doi.org/10.1145/3359294

[45] Dan M Kahan. 2017. Misconceptions, misinformation, and the logic of identity-protective cognition. (2017).

[46] Ben Kaiser, Jerry Wei, Eli Lucherini, Kevin Lee, J. Nathan Matias, and Jonathan Mayer. 2021. Adapting Security Warnings to Counter Online Disinformation. In *30th USENIX Security Symposium (USENIX Security 21).* USENIX Association, 1163–1180. https://www.usenix.org/conference/usenixsecurity21/presentation/kaiser

[47] Cecilia King. 2016. Fake News Onslaught Targets Pizzeria as Nest of Child-Trafficking. (Nov 2016). https://www.nytimes.com/2016/11/21/technology/fact-check-this-pizzeria-is-not-a-child-trafficking-site.html

[48] Jan Kirchner and Christian Reuter. 2020. Countering Fake News: A Comparison of Possible Solutions Regarding User Acceptance and Effectiveness. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2 (oct 2020).

[49] Ben Kirman, Conor Lineham, and Shaun Lawson. 2012. Exploring Mischief and Mayhem in Social Computing or: How We Learned to Stop Worrying and Love the Trolls. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI EA '12).* Association for Computing Machinery, New York, NY, USA, 121–130. https://doi.org/10.1145/2212776.2212790

[50] David M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, and Jonathan L. Zittrain. 2018. The science of fake news. *Science* 359, 6380 (2018), 1094–1096. https://doi.org/10.1126/science.aao2998 arXiv:https://www.science.org/doi/pdf/10.1126/science.aao2998

[51] Stephan Lewandowsky and Sander van der Linden. 2021. Countering Misinformation and Fake News Through Inoculation and Prebunking. *European Review of Social Psychology* 32, 2 (2021), 348–384.

[52] Xingyu Liu, Patrick Carrington, Xiang 'Anthony' Chen, and Amy Pavel. 2021. What Makes Videos Accessible to Blind and Visually Impaired People?. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21).* Association for Computing Machinery, New York, NY, USA, Article 272, 14 pages. https://doi.org/10.1145/3411764.3445233

[53] Clare Llewellyn, Laura Cram, Adrian Favero, and Robin L. Hill. 2018. Russian Troll Hunting in a Brexit Twitter Archive. In *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries* (Fort Worth, Texas, USA) *(JCDL '18).* Association for Computing Machinery, New York, NY, USA, 361–362. https://doi.org/10.1145/3197026.3203876

[54] Haley MacLeod, Cynthia L Bennett, Meredith Ringel Morris, and Edward Cutrell. 2017. Understanding blind people's experiences with computer-generated captions of social media images. In *proceedings of the 2017 CHI conference on human factors in computing systems.* 5988–5999.

[55] Cameron Martel, Gordon Pennycook, and David G Rand. 2020. Reliance on emotion promotes belief in fake news. *Cognitive research: principles and implications* 5 (2020), 1–20.

[56] Lee McIntyre. 2018. *Post-truth.* MIt Press.

[57] Meta. 2022. How We're Tackling Misinformation Across Our Apps. https://about.fb.com/news/2021/03/how-were-tackling-misinformation-across-our-apps/

[58] Meredith Ringel Morris, Annuska Zolyomi, Catherine Yao, Sina Bahram, Jeffrey P. Bigham, and Shaun K. Kane. 2016. "With Most of It Being Pictures Now, I Rarely Use It": Understanding Twitter's Evolving Accessibility to Blind Users. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16).* Association for Computing Machinery, New York, NY, USA, 5506–5516. https://doi.org/10.1145/2858036.2858116

[59] Garrett Morrow, Briony Swire-Thompson, Jessica Polny, Matthew Kopec, and John Wihbey. 2021. The Emerging Science of Content Labeling: Contextualizing Social Media Content Moderation. (2021). https://doi.org/10.2139/ssrn.3742120

[60] Shuo Niu, Jaime Garcia, Summayah Waseem, and Li Liu. 2022. Investigating How People with Disabilities Disclose Difficulties on YouTube. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility.* 1–5.

[61] Brendan Nyhan and Jason Reifler. 2010. When corrections fail: The persistence of political misperceptions. *Political Behavior* 32, 2 (2010), 303–330.

[62] Orestis Papakyriakopoulos and Ellen Goodman. 2022. The Impact of Twitter Labels on Misinformation Spread and User Engagement: Lessons from Trump's Election Tweets. In *Proceedings of the ACM Web Conference 2022.* 2541–2551.

[63] Gordon Pennycook, Jonathon McPhetres, Yunhao Zhang, Jackson G. Lu, and David G. Rand. 2020. Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. *Psychological*

Science 31, 7 (2020), 770–780.

[64] Gordon Pennycook and David G. Rand. 2020. Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality* 88, 2 (2020), 185–200. https://doi.org/10.1111/jopy.12476 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/jopy.12476

[65] Gordon Pennycook and David G. Rand. 2021. The Psychology of Fake News. *Trends in Cognitive Sciences* 25, 5 (2021), 388–402.

[66] Julie Posetti and Alice Matthews. 2018. A short guide to the history of 'fake news' and disinformation. *International Center for Journalists* 7, 2018 (2018), 2018–07.

[67] Man pui Sally Chan, Christopher R. Jones, Kathleen Hall Jamieson, and Dolores Albarracín. 2017. Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation. *Psychological Science* 28, 11 (2017), 1531–1546.

[68] Richard Rogers. 2020. Deplatforming: Following extreme Internet celebrities to Telegram and alternative social media. *European Journal of Communication* 35, 3 (2020), 213–229.

[69] Mohammad Hammas Saeed, Shiza Ali, Jeremy Blackburn, Emiliano De Cristofaro, Savvas Zannettou, and Gianluca Stringhini. 2022. TrollMagnifier: Detecting State-Sponsored Troll Accounts on Reddit. In *2022 IEEE Symposium on Security and Privacy (SP)*. 2161–2175. https://doi.org/10.1109/SP46214.2022.9833706

[70] Emily Saltz, Claire R Leibowicz, and Claire Wardle. 2021. *Encounters with Visual Misinformation and Labels Across Platforms: An Interview and Diary Study to Inform Ecosystem Approaches to Misinformation Interventions.* Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411763.3451807

[71] Laura Savolainen. 2022. The shadow banning controversy: perceived governance and algorithmic folklore. *Media, Culture & Society* 44, 6 (2022), 1091–1109.

[72] Filipo Sharevski, Raniem Alsaadi, Peter Jachim, and Emma Pieroni. 2022. Misinformation warnings: Twitter's soft moderation effects on COVID-19 vaccine belief echoes. *Computers & Security* 114 (2022), 102577. https://doi.org/10.1016/j.cose.2021.102577.

[73] Filipo Sharevski, Raniem Alsaadi, Peter Jachim, and Emma Pieroni. 2022. Misinformation warnings: Twitter's soft moderation effects on COVID-19 vaccine belief echoes. *Computers & Security* 114 (2022), 102577.

[74] Filipo Sharevski, Amy Devine, Peter Jachim, and Emma Pieroni. 2022. "Gettr-ing" User Insights from the Social Network Gettr. https://truthandtrustonline.com/wp-content/uploads/2022/10/TTO_2022_proceedings.pdf.

[75] Filipo Sharevski, Amy Devine, Peter Jachim, and Emma Pieroni. 2022. Meaningful Context, a Red Flag, or Both? Preferences for Enhanced Misinformation Warnings Among US Twitter Users. In *Proceedings of the 2022 European Symposium on Usable Security* (Karlsruhe, Germany) *(EuroUSEC '22)*. Association for Computing Machinery, New York, NY, USA, 189–201. https://doi.org/10.1145/3549015.3555671 https://doi.org/10.1145/3549015.3555671.

[76] Filipo Sharevski, Amy Devine, Emma Pieroni, and Peter Jachim. 2023. Folk Models of Misinformation On Social Media. In *Network and distributed system security symposium.* https://dx.doi.org/10.14722/ndss.2023.24293.

[77] Filipo Sharevski, Alice Huff, Peter Jachim, and Emma Pieroni. 2022. (Mis)perceptions and engagement on Twitter: COVID-19 vaccine rumors on efficacy and mass immunization effort. *International Journal of Information Management Data Insights* 2, 1 (2022), 100059.

[78] Filipo Sharevski, Peter Jachim, Emma Pieroni, and Nate Jachim. 2021. VoxPop: An Experimental Social Media Platform for Calibrated (Mis)Information Discourse. In *New Security Paradigms Workshop* (Virtual Event, USA) *(NSPW '21)*. Association for Computing Machinery, New York, NY, USA, 88–107. https://doi.org/10.1145/3498891.3498893

[79] Filipo Sharevski and Benjamin Kessell. 2023. Fight Fire with Fire: Hacktivists' Take on Social Media Misinformation. In *Nineteenth Symposium on Usable Privacy and Security (SOUPS 2023)*. USENIX Association, Anaheim, CA, 19–36. https://www.usenix.org/conference/soups2023/presentation/sharevski

[80] Filipo Sharevski, Jennifer Vander Loop, Peter Jachim, Amy Devine, and Emma Pieroni. 2023. Talking Abortion (Mis) information with ChatGPT on TikTok. (2023). https://arxiv.org/abs/2303.13524.

[81] Filipo Sharevski, Paige Treebridge, and Jessica Westbrook. 2019. Manipulation of Perceived Politeness in a Web-Based Email Discourse through a Malicious Browser Extension. In *Proceedings of the New Security Paradigms Workshop* (San Carlos, Costa Rica) *(NSPW '19)*. Association for Computing Machinery, New York, NY, USA, 31–41. https://doi.org/10.1145/3368860.3368863

[82] Peter Warren Singer and Emerson T Brooking. 2018. *LikeWar: The weaponization of social media.* Eamon Dolan Books.

[83] Mohit Singhal, Chen Ling, Nihal Kumarswamy, Gianluca Stringhini, and Shirin Nilizadeh. 2022. SoK: Content moderation in social media, from guidelines to enforcement, and research to practice. *arXiv preprint arXiv:2206.14855* (2022).

[84] Alexa F Siu, Danyang Fan, Gene SH Kim, Hrishikesh V Rao, Xavier Vazquez, Sile O'Modhrain, and Sean Follmer. 2021. COVID-19 highlights the issues facing blind and visually impaired people in accessing data on the web. In *Proceedings of the 18th International Web for All Conference*. 1–15.

[85] Jeff Smith. 2017. Designing Against Misinformation. *Medium* (2017). https://medium.com/facebook-design/designing-against-misinformation-e5846b3a1e2.

[86] Social Security Administration. 2023. 404.1581. Meaning of blindness as defined in the law. https://www.ssa.gov/OP_Home/cfr20/404/404-1581.htm

[87] Abigale Stangl, Meredith Ringel Morris, and Danna Gurari. 2020. " Person, Shoes, Tree. Is the Person Naked?" What People with Vision Impairments Want in Image Descriptions. In *Proceedings of the 2020 chi conference on human factors in computing systems*. 1–13.

[88] Briony Swire-Thompson, John Cook, Lucy H. Butler, Jasmyne A. Sanderson, Stephan Lewandowsky, and Ullrich K. H. Ecker. 2021. Correction format has a limited role when debunking misinformation. *Cognitive Research: Principles and Implications* 6, 1 (2021), 83.

[89] Briony Swire-Thompson, Joseph DeGutis, and David Lazer. 2020. Searching for the Backfire Effect: Measurement and Design Considerations. *Journal of Applied Research in Memory and Cognition* 9, 3 (2020), 286–299. https://doi.org/10.1016/j.jarmac.2020.06.006

[90] Emily Thorson. 2016. Belief echoes: The persistent effects of corrected misinformation. *Political Communication* 33, 3 (2016), 460–480.

[91] TikTok. 2023. TikTok Safety. https://www.tiktok.com/safety/en-us/topics/.

[92] Tollefson, Jeff. 2023. Disinformation researchers under investigation: what's happening and why. https://www.nature.com/articles/d41586-023-02195-3

[93] Twitter. 2020. Information Operations. https://transparency.twitter.com/en/reports/information-operations.html

[94] Twitter. 2021. Birdwatch: A community-driven approach to address misinformation on Twitter. https://twitter.github.io/birdwatch/about/overview/

[95] Jay J Van Bavel and Andrea Pereira. 2018. The partisan brain: An identity-based model of political belief. *Trends in cognitive sciences* 22, 3 (2018), 213–224.

[96] Luis Vargas, Patrick Emami, and Patrick Traynor. 2020. On the Detection of Disinformation Campaign Activity with Network Analysis. In *Proceedings of the 2020 ACM SIGSAC Conference on Cloud Computing Security Workshop* (Virtual Event, USA) *(CCSW'20)*. Association for Computing Machinery, New York, NY, USA, 133–146. https://doi.org/10.1145/3411495.3421363

[97] Emily K. Vraga and Leticia Bode. 2020. Defining Misinformation and Understanding its Bounded Nature: Using Expertise and Evidence for Describing Misinformation. *Political Communication* 37, 1 (2020), 136–144.

[98] Nathan Walter, Jonathan Cohen, R. Lance Holbert, and Yasmin Morag. 2020. Fact-Checking: A Meta-Analysis of What Works and for Whom. *Political Communication* 37, 3 (05 2020), 350–375.

[99] Sarah Myers West. 2018. Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society* 20, 11 (2018), 4366–4383. https://doi.org/10.1177/1461444818773059 arXiv:https://doi.org/10.1177/1461444818773059

[100] Gill Whitney and Irena Kolar. 2020. Am I missing something? *Universal Access in the Information Society* 19, 2 (2020), 461–469.

[101] Thomas Wood and Ethan Porter. 2019. The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior* 41, 1 (2019), 135–163.

[102] Samuel C Woolley and Philip N Howard. 2018. *Computational propaganda: Political parties, politicians, and political manipulation on social media.* Oxford University Press.

[103] Shaomei Wu and Lada A Adamic. 2014. Visually impaired users on an online social network. In *Proceedings of the sigchi conference on human factors in computing systems*. 3133–3142.

[104] Yahoo! News. 2023. With scientists split between the lab-leak theory or a zoonotic origin, experts say the true origin of the pandemic may take years to be known — if ever. https://www.tiktok.com/@yahoonews/video/7205328311835282734

[105] YouTube. 2023. COVID-19 medical misinformation policy. https://support.google.com/youtube/answer/9891785?hl=en

[106] YouTube. 2023. Misinformation policies. https://support.google.com/youtube/answer/10834785?hl=en

[107] Savvas Zannettou. 2021. "I Won the Election": An Empirical Analysis of Soft Moderation Interventions on Twitter. *Proceedings of the International AAAI Conference on Web and Social Media* 15, 1 (May 2021), 865–876.

[108] Savvas Zannettou, Barry Bradlyn, Emiliano De Cristofaro, Haewoon Kwak, Michael Sirivianos, Gianluca Stringini, and Jeremy Blackburn. 2018. What is Gab: A Bastion of Free Speech or an Alt-Right Echo Chamber. In *Companion Proceedings of the The Web Conference 2018* (Lyon, France) *(WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1007–1014. https://doi.org/10.1145/3184558.3191531

[109] Savvas Zannettou, Tristan Caulfield, Emiliano De Cristofaro, Nicolas Kourtelris, Ilias Leontiadis, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. 2017. The Web Centipede: Understanding How Web Communities Influence Each Other through the Lens of Mainstream and Alternative News Sources. In *Proceedings of the 2017 Internet Measurement Conference* (London, United Kingdom) *(IMC '17)*. Association for Computing Machinery, New York, NY, USA, 405–417. https://doi.org/10.1145/3131365.3131390

[110] Savvas Zannettou, Tristan Caulfield, William Setzer, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. 2019. Who Let The Trolls Out?

Towards Understanding State-Sponsored Trolls. In *Proceedings of the 10th ACM Conference on Web Science* (Boston, Massachusetts, USA) *(WebSci '19)*. Association for Computing Machinery, New York, NY, USA, 353–362. https://doi.org/10.1145/3292522.3326016

[111] Mingrui Ray Zhang, Mingyuan Zhong, and Jacob O Wobbrock. 2022. Ga11y: An Automated GIF Annotation System for Visually Impaired Users. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–16.

APPENDIX

Interview Questions

(1) What social media platforms do you use the most?
   - Facebook
   - TikTok
   - YouTube
   - Other (Please specify)

(2) What device do you prefer to consume your social media on?
   - Phone
   - Tablet
   - Laptop/desktop
   - Other (please specify)

(3) What assistive technology do you use when consuming social media (select all that apply)
   - Screen reader (VoiceOver, NVDA, JAWS, Narrator, TalkBack, etc.)
   - Screen magnification
   - Color inversion
   - Other (Please specify)

(4) We are sending you a link for a piece of content from your social media platform of choice. Please provide your opinion on the content's accuracy and reliability. Note anything that stands out to you. Feel free to vocalize your thought process. Take the time you need to review this content.

(5) Did you notice any platform-provided warnings and/or recommendations about the content you viewed?

(6) How did these warnings and/or recommendations affect your perception of the content's accuracy and reliability?

(7) Have you encountered such warnings and/or recommendations on other platforms [which ones?] and on what content [list content]?

(8) What is your opinion on how this and other social media platforms handle misleading content, deemed potentially as misinformation? Consider their utilization of warnings and/or recommendations.

(9) What is your opinion on how these warnings and/or recommendations are made accessible (or not) for blind people or people with visual impairments?

(10) What is your opinion on how misinformation warnings and/or recommendations could/should be made adequately accessible for blind people or people with visual impairments?

(11) Anything else you want to add on this topic or your experience with accessibility, misinformation, warnings, and social media?

(12) Which statement best describes you
   - I am totally blind and am unable to perceive lights and shapes
   - I am blind and am able to perceive lights and/or shapes
   - I am low vision and consider myself to be legally blind
   - I am low vision but I do not consider myself to be legally blind

(13) Demographics: Age, race/ethnicity, gender, visual diagnosis, education level, political self-identification

Codebook

- **Misinformation Warning Label** Codes related to the direct interaction with the misinformation warning labels assigned to the social media posts. The category is divided into three sub-categories.
  - **Direct access** The sub-category collects codes pertaining to the way the warning label is located and accessed in the user interface.
    - **Without difficulties** The participant was able to locate the warning label without any difficulties (i.e. their assistive technology made a sufficient friction for them to hear or see the label's text)
    - **Without difficulties and extrapolated their past *accessibility* experience** The participant was able to locate the warning label without any difficulties but also commented how the particular label fares with their past general accessibility experience with application notifications
    - **Without difficulties and extrapolated their past experience with *warning labels*** The participant was able to locate the warning label without any difficulties but also commented how the particular label fares with their past general experience with labels on social media
    - **With difficulties** The participant was able to locate the warning label but with difficulties due to the confusing text-to-speech translation of the label's text by the VoiceOver assistant
  - **Point-back access** The sub-category collects codes pertaining to the way the warning label is located and accessed in the user interface, after the researchers pointed the participants back to it when they initially missed it.
    - **Without difficulties** The participant was able to return back and locate the warning label without any difficulties (i.e. their assistive technology made a sufficient friction for them to hear or see the label's text)
    - **Without difficulties and extrapolated their past *accessibility* experience** The participant was able to to return back and locate the warning label without any difficulties but also commented how the particular label fares with their past general accessibility experience with application notifications
    - **With difficulties** The participant was able to locate the warning label but with difficulties due to the confusing text-to-speech translation of the label's text by the VoiceOver assistant
  - **Accessibility opinion** The sub-category collects codes pertaining to the general opinion regarding the accessibility of the misinformation warning labels amoung the participants

    - **Accessible** The participant is on the opinion that the warning labels are accessible through any of the assistive technologies (e.g. screen readers, magnification, or contrast)
    - **Accessible, but without utility** The participant is on the opinion that the warning labels are accessible, but they are practically useless as they are easily ignored or are unintelligible, rendering them *de facto* inaccessible
    - **Not accessible** The participant is on the opinion that the warning labels are not accessible through any of the assistive technologies (e.g. screen readers, magnification, or contrast)
- **Truth Discernment** Codes related to the way the misinformation warning labels helped the participants with discerning the truth about the statements in the social media posts. The category is divided into two sub-categories.
  - **Truth Discernment** The sub-category collects codes pertaining to the way the warning label helped the participants with the truth discernment relative to the statement in the social media post.
    - **Helped** The misinformation warning label did help with the truth discernment
    - **Did not help** The misinformation warning label did not help with the truth discernment
    - **Did not help and invoked commentary** The misinformation warning label did not help with the truth discernment and invoked a commentary from the participant about the truth discernment effect of misinformation warnings on social media
  - **General Opinion** The sub-category collects codes pertaining to the way general opinion the participants provided regarding the intended effect of the warning labels relative to truth discernment.
    - **Labels are in the way** The participant opines that the labels are interrupt the flow of accessing content on the platform
    - **Labels are intrusive** The participant opines that the labels are intrusive and promote the platform's "hidden" agenda for interfering with the free speech of the users
    - **Labels are wrongly substantiated** The participant opines that the labels are wrongly substantiated to the posts, i.e. the posts are not misinformation and do not need to be labeled as such
    - **Labels are indistinguishable** The participant opines that the labels are indistinguishable and therefore the labels don't help with the truth discernment
    - **Labels are useful** The participant opines that the labels provide a bit of context about the content they are substantiated to
- **Accessibility Redesign Recommendations** Codes related to the recommendations provided by the participants for improving both the accessibility and the truth discernment of the misinformation warning labels
  - **Labels entirely removed** The participant want the labels entirely removed from the platforms
  - **Prefers a verbose cover** The participant prefers a verbose cover instead of a misinformation warning label
  - **Prefers a verbose before-content label** The participant prefers a verbose label that is substantiated before the content on the screen
  - **Prefers a before-content audio signal or vibration in addition to a cover or a label** The participant prefers an audio signal or vibration that is substantiated before the content on the screen in addition to the warning label
  - **Prefers a stark contrast, bold/large font, standout colors different from the platforms' aesthetics** The participant prefers stark contrast, bold/large font, standout colors different from the platforms' aesthetics for formating of the misinformation warning labels