The Burden of Bearing Witness: Digital Practices of Marginalized Social Media Users in High-Stakes Contexts

Sena A. Kojah
School of Information
University of Michigan
Ann Arbor, Michigan, USA
senami@umich.edu

Hibby Thach
School of Information
University of Michigan
Ann Arbor, Michigan, USA
hibby@umich.edu

Oliver L. Haimson
School of Information
University of Michigan
Ann Arbor, Michigan, USA
haimson@umich.edu

Michaelanne Thomas
School of Information
University of Michigan
Ann Arbor, Michigan, USA
mmtd@umich.edu

Abstract

Social media platforms and their governance policies often fail marginalized users in high-stakes contexts, including war, violent attacks, human rights violations, humanitarian crises, and situations of systemic oppression. Through interviews, autoethnography, and digital ethnography, this paper presents three case studies from Venezuela, Nigeria, and the United States to examine how marginalized populations engage with social media in non-normative ways. We analyze how platform design and policies intersect with participants' identities, marginalization, and labor. Our central finding is that users' urgent infrastructural and contextual needs are often overlooked, revealing structural flaws in social media design that mimic physical-world power asymmetries. In response, users develop innovative workarounds, engage in self-censorship, and adopt coping strategies, undertaking additional, often invisible, sociotechnical repair work that reinforces their precarity. To address these complex needs, we urge social media companies to collaborate with marginalized users to integrate alternative infrastructural features, such as emergency response tools and exit mechanisms for wellbeing.

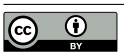
CCS Concepts

• **Human-centered computing** → **Social media**; **Collaborative and social computing**; **Collaborative and social computing theory, concepts and paradigms**;

Keywords

Social media, repair work, alternative infrastructural design, crisis, high-stakes contexts, content moderation, marginality, precarity

ACM Reference Format:
Sena A. Kojah, Hibby Thach, Oliver L. Haimson, and Michaelanne Thomas. 2026. The Burden of Bearing Witness: Digital Practices of Marginalized Social Media Users in High-Stakes Contexts. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3772318.3791924



ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3791924

1 Introduction

Social media platforms and platform governance policies fail in high-stakes contexts like war, violence, disasters, humanitarian crises, harassment, displacement, and other oppressive conditions [48]. Yet despite these breakdowns, in moments of heightened precarity, people's need to share and receive critical information increases [29, 29, 48, 56, 79]. Thus, for marginalized and underserved populations facing infrastructural breakdown, and in moments of crisis, social media can serve as an alternative infrastructure to support communal bonds and facilitate critical resource management [29, 55]. People in high-stakes contexts often rely on social media platforms to document their experiences for internal and external audiences [63]. Yet when they do, they contend with platform features and policies that complicate their ability to use social media in ways that respond to their contextual needs.

To understand the complexities that manifest when marginalized users in high-stakes contexts use social media to document their experiences, and to examine how these people's identities, marginality, labor, and social media use intersect, we asked the following research questions:

RQ1. How do marginalized populations in high-stakes contexts use social media in ways that fall outside of normative social media use?

RQ2. How do platform features and social media use intersect with identity, marginality, and labor in high-stakes contexts?

RQ3. How do marginalized populations in high-stakes contexts negotiate tensions within platform features?

We present three case studies – diasporic groups coping with the humanitarian crisis in Venezuela, ethnic minorities facing violence and human rights violations in Nigeria, and trans people creating technology while facing oppression in the United States – to examine how marginalized populations in high-stakes contexts experience social media platform features and policies. Marginalized populations are those who are excluded from participation in technological systems due to a lack of access based on their gender, sexuality, socio-economic status, ethnicity, and other factors [59].

As Liang et al discuss, intention around the use of "marginalized" matters, as terms like minoritized, underserved, or vulnerable all have different meanings, often revolving around society forgetting certain groups of people or labeling them as weak [28, 98]. In the context of this study, marginalized populations are people whose experiences with and access to social digital technologies, especially in moments of precarity, are limited based on their identities and geographic locations. We specifically focus on groups marginalized due to their nationality, ethnicity, and/or gender. Precarity in the context of this study refers to all forms of algorithmic uncertainty that social media users in high-stakes contexts experience which is complicated by their identity and marginality.

This paper expands previous research examining the infrastructural care efforts that social media users engage in during moments of crisis [29]. From prior work, some examples of how crisis intersects with social technologies include uncertainties around platform governance policies during social media use in conflict situations [55, 80], marginalized populations' need for digital repositories on social media to document violence and human rights violations [55], and the disproportionate ways that online platforms moderate gender and bodies [42, 56].

In all three case studies, our findings show that participants relied on social media to post content related to their individual and/or communal survival. However, they had to navigate design features within platforms (e.g Facebook, X, Discord) in ways that intersected negatively with their identities, increased the labor they had to perform, and deepened their marginality. Meanwhile, social media platforms de-prioritized participants' high-stakes infrastructural and contextual needs. We also found that complex content moderation decisions about issues like misinformation, graphic images from people in conflict zones, and content related to trans identity and bodies did not receive the specialized attention they required. While participants in all case studies experienced forms of platform governance, the context in which they experienced it and how platform features interacted with their experiences differed. In each of these contexts, unclear platform design led to additional labor for participants engaging in infrastructural care efforts using social media platforms.

Platforms with design features that do not accommodate peripheral social media use in high-stakes contexts are like buildings in need of repair work. Yet complexity arises because these buildings are inhabited by people and communities (site users, designers, and employees), each with unique needs and motivations. The participants in all three vignettes scaffold their lived experiences and identity onto these platforms in unconventional ways that challenge existing design. We find that platform failures limits how marginalized people can use social media platforms in moments of heightened precarity and high-stakes contexts. This creates a stakeholder class that is further plunged into precarity and marginality. The experiences of these social media users reveal the structural flaws of platforms, which is symptomatic of platforms' inability to accommodate all types of users.

Our findings contribute new ways of understanding how social media users in high-stakes contexts have created, and continue to develop, complex solutions, alternative infrastructures, and ways of circumventing platforms' failures to account for their complex needs. These alternative ways of navigating platforms differ from typical social media use, as everyday users do not have to contend with these crisis contexts. To address the breakdown of social media for marginalized users in high-stakes contexts and create more inclusive platforms, we recommend that social media platforms collaborate with marginalized users to integrate alternative infrastructural features, such as emergency response tools and exit mechanisms for wellbeing, that these groups are in some cases already using. Simultaneously, we recognize social media companies have little incentive to incorporate such design recommendations and continue to neglect the needs of their most vulnerable stakeholders. We grapple with these paradoxes throughout this paper.

2 Related Work

2.1 Social Media Platforms as Sites for Coloniality in Social Computing

In many countries worldwide and for populations on the margins of society, sociotechnical systems reinstate colonial structures that hinder individual expression [19]. Hierarchical categorizations introduced by colonizers into non-Western societies continue to define and shape how people interact within communities and how those interactions are translated into computing [20, 87]. On social media platforms, coloniality consolidates existing material and colonial relations when diverse and marginalized populations interact online, thereby reinforcing global power asymmetries [83]. The dominance of digital technology from the Global North in the lives of people in the Global South means that they are subject to the power and extractive capabilities within socio-technical systems [57]. However, that power is not shared with them, which contributes to their further subjugation and invisibility. Further, the design of social media platforms and their policies make it easy for platforms to insert themselves into the local socio-political and economic conditions of marginalized communities and, in the process, reinforce racialized and social-political oppression in collaboration with the state and dominant ethnic and racial groups [83].Research shows that colonial framings of identity embedded within content moderation systems can misunderstand non-normative content as negative and lead to automated content moderation [21]. This leads marginalized populations to worry about content removals that further marginalize them [21]. Coloniality also manifests as the extraction of information and data from the epistemic architectures of diverse and marginal populations as a way to maintain digital ecologies [100] with platforms doing little to no work to acknowledge the diversity of being outside of Western hegemony [87]. This coloniality translates into the way that platform governance policies perform at the intersection of marginality and labor.

2.2 The Design of Platform Governance and Marginality

Social media companies struggle with understanding and implementing how best to moderate content from and about marginalized and diverse populations across the world [6, 23, 42, 61, 73, 87]. Current content moderation technologies and policies have yet to come to terms with the complexities of human interaction [24]. Social media users with marginalized identities disproportionately experience content removals and opaque platform governance policies

[6, 23, 42]. The wide range of online content produced by and about diverse populations means that broad-stroke moderation policies often lead to disastrous outcomes for those on the margins of society, such as Black people and Global South populations [45, 87]. For example, content that is flagged as violent, such as a beheading, might be political and newsworthy, while content posted as comedy could be misogynistic, transphobic, or homophobic [13]. Additionally, bad faith actors exploit the anonymity of the internet to target political opponents [13]. When moderating a highly diverse set of users, policies cannot be applied across all contexts universally, and instead appear impractical and opaque [9, 42].

There are many forms of diversity and marginality where applying one-size-fits-all content moderation fails, but for the purpose of this paper, we specifically examine political marginality, ethnic diversity, and gender diversity. Trans people are often subjected to disproportionate content moderation because of binary or static views about gender that influence content moderation policy and implementation across popular social media platforms [42, 65, 78, 96]. These removals occur when trans people are posting about their experiences, their bodies, seeking healthcare information, or simply participating in public discourse [3, 40, 65, 96]. To counter disproportionate removals and transphobic abuse, many trans creators resort to performing additional labor by moderating/filtering their user bases, curating their content in ways that stifle their individuality and self-expression, and relying on community governance [4, 96]. Others invent language using "algospeak" to self-censor and circumvent words that they perceive algorithms will remove or suppress [91]. These experiences share a similar thread with what researchers have observed about how social media companies moderate populations in the Global South, as we describe next.

The unique aspects of moderating populations interacting on social media in the Global South lie in the complexity and diversity of human experiences that are not accounted for in content moderation policies predominantly emerging from the Western world. The scant knowledge about how users in the Global South interact with opaque moderation systems reveals the perpetuation of colonial power relations that center Western norms around acceptable speech, content, expressions of self, and notions of identity and gender [87]. On the other hand, platforms often fail to moderate fake news and incitements of violence targeting marginalized people, leading to harm [22]. Ethnic and religious minorities in the Global South bear a significant portion of the devastating consequences of disproportionate content moderation [82, 92]. Internet and social media use continues to soar in countries in the Global South [77], yet, social media companies prioritize English-speaking populations, leading to dangerous outcomes for the rest of the world [85]. This means that hate and targeted abuse in countries with low-resourced languages spread faster because there is no way of detecting how harmful they are [74]. It also means that social media users in these countries must navigate harmful landmines when they perform mundane tasks on social media [70]. One study found that YouTube exposed Amharic-speaking users to sexual content when they made harmless searches for things like Amharic religious music [70]. Other harms resulting from social media platforms' failure to account for Global South populations include the exacerbation of violent ethno-religious conflicts in places like Sri Lanka [92], Myanmar [68, 82], India [33] and Ethiopia [62, 92]. During active conflicts in non-Western contexts, content moderation policies and enforcement typically worsen situations, partly because of poor transparency, often leading to loss of life and property [62, 82].

Critical groups, including activists, journalists, and minoritized communities, are increasingly using social media to document human rights violations and conflicts in graphic and controversial ways [6, 63]. However, this content is subject to removal by platforms applying one-size-fit-all content moderation decisions that do not consider context-specific situations [6].

Many social media users often perceive content moderation policies, decisions, and procedures as arbitrary. Content that might be harmful to some users might be considered useful to others. Social media users who are arbitrarily and disproportionately moderated place the blame squarely on human and algorithmic content moderation's affordances, or lack thereof [5]. Algorithmic decision-making often does not translate to users' traditional perceptions of justice and fairness [8, 9]. Research suggests that users believe algorithmic content moderation and decision-making systems are impersonal and lack the ability to contextualize and articulate why and how they make decisions, and why they consider those decisions fair [5, 8, 45]. Algorithmic content moderation exacerbates several other challenges, including: obscuring the inherently political nature of large-scale speech regulation [38]; enforcing punitive opacity that requires content creators to perform algorithmic labor (often leading to economic precarity) [60]; reinforcing power imbalances and inconsistent appeals processes [97]; perpetuating biases from human raters who provide training data; and amplifying subjective norms that further obscure already non-transparent content moderation policies and practices [9].

Many often tout that human content moderation is the solution to the mechanical administration of moderation governance by algorithms. However, research shows humans also struggle to navigate human complexities [35, 67]. Still, human content moderators are essential if they can provide more holistic content moderation that algorithmic systems often fail to achieve [35, 67].

2.3 Invisibilized Digital Labor

Invisible labor refers to essential work that is systematically undervalued and obscured from view [17, 18]. In digital economies, this labor takes many forms, including hidden, affective, and algorithmically mediated practices that sustain platforms and creator industries [54, 72]. One major area of this literature focuses on how cultural workers navigate opaque algorithmic systems that determine both visibility and income. Existing research [27, 56] ascribes this to algorithmic precarity, where workers' livelihoods depend on unpredictable recommendation systems. Extending this, Duffy and Meisner [26] show how platform governance practices intensify marginalized creators' struggles with algorithmic (in)visibility, forcing them to constantly adapt their content and labor routines to remain discoverable. Research on identity-based platform use [39, 88] demonstrates how users negotiate with algorithms to sustain control over their visibility, while Duffy and Sawey [27] showcase the paradox of creators who are hyper-visible as brands, yet invisible as workers. Much of this hidden labor is unpaid, from mitigating shadowbans [56] to maintaining branded personas.

Algorithms thus reinforce precarity and platform dependency as they become increasingly embedded in digital economies. Glatt [37] shows how influencers diversify labor across multiple platforms to manage instability, yet this strategy further amplifies hidden work as creators juggle platform cultural norms and technical demands. Constant self-branding and the circulation of "algorithmic gossip," or informal knowledge, rumors, and speculative advice on managing visibility and access within platform algorithms, are central to sustaining reach [10, 32]. Such practices include tracking engagement [7] and monitoring what Glatt [37] calls "popularity metrics." Despite this precarity, digital laborers are encouraged to self-exploit through discourses of "empowerment" and affective labor [7, 72]. Casilli [16] urges scholars to approach these dynamics globally and to take a decolonial approach, noting that digital economies rely on low-paid, hidden work that often crosses national borders.

Non-Western research further highlights these low-paid, hidden work contexts. Studies of gig workers in India and Indonesia show how waiting time, repetitive digital interactions, and algorithmic opacity contribute to hidden forms of labor that undermine the promise of "flexibility" [94, 99]. In the Philippines, "digital labor bayanihan" captures how workers build informal infrastructures of mutual aid through guidance and affective care often unrecognized by platforms [89]. Outsourced content moderation, like labor outsourced to the Philippines, likewise illustrates how emotionally taxing, stigmatized work remains invisible within global platform economies [81]. Similarly, in Sub-Saharan Africa, workers face barriers of algorithmic invisibility and rating dependence that obscure their contributions to digital economies [25].

This invisibility extends to moderation systems themselves. Gillespie [36] highlights how moderation tools like "do not recommend" invisibly govern and impact creators' reach and income. Users often engage in moderation and repair practices themselves that remain hidden. For example, Thach et al [95] detail how marginalized Twitch and Reddit users engage in invisible and often unpaid moderation practices to navigate harassment and toxic technocultures, or anti-progressive platform cultures that enable harm towards marginalized users [64]. Similarly, crowd workers must often fight to make their hidden labor visible to their employers and the public [50]. Users also create "folk theories," or "intuitive, causal explanations about a system," [31] and despite having little real insight into platform systems, these folk theories influence their everyday practices around algorithms [30, 34, 66].

We contribute to the literature on invisible digital labor by examining how three case studies of users in crisis navigate substantial work maneuvering around algorithmic invisibility. Additionally, we expand the current work on infrastructure and repair work to include these case studies of users experiencing heightened levels of precarity in different global contexts.

3 Methods

We present three qualitative and ethnographic case studies written by the first, second, and fourth authors (respectively), drawn from data collected by all authors separately as part of three studies. Due to their commonalities and unique aspects, we chose to combine our case studies, documenting content moderation experiences of marginalized and minoritized populations in the Global South and the Global North. Our study contexts are ethnic minorities in Nigeria facing violence and human rights violations, diasporic groups during the humanitarian crisis in Venezuela, and trans people facing oppression in the U.S. In each of our case studies, we spoke with people who experienced or witnessed violence that they wished to share online, used social media as a form of activism and documentation, and experienced online harassment and violence. Our engagement with qualitative and ethnographic methods enabled us to observe how people make social meaning from their experiences [12, 46]. Using data from interviews, autoethnography, and digital ethnography, our case studies make connections between the distinct yet overlapping needs of marginalized and minoritized populations in high-stakes contexts when they interact online with content that often falls outside of the realm of traditional content moderation.

In the first case study, the fourth author moved between physical and virtual field sites, where the reality and dynamic nature of Venezuelan content creators' experiences were translated into physical, mental, and emotional fatigue. The original project, conducted in 2018, aimed to understand how members of the Venezuelan diaspora coped with the ongoing humanitarian crisis and documented their sociotechnical infrastructural actions to support family and friends at a distance. The physical field sites included an initial recruitment focus in a major New England city in the U.S., where some interviews were conducted in person at locations such as student centers, though the study quickly expanded to virtual sites and phone/Skype interviews with participants across the U.S., Chile, and Argentina. The case study draws on data from 13 in-depth interviews with participants born in Venezuela who had since migrated to the U.S., Chile, and Argentina, respectively. It also included eight months of virtual participant observation across Facebook, Instagram, WhatsApp, and X (formerly Twitter), during which the fourth author observed, participated, and took field notes regarding the various strategies used to solicit support and discuss the current situation in Venezuela. The fourth author is a white American woman who spent her early childhood living in countries where Spanish is the primary language. She is fluent in both English and Spanish, having learned both languages simultaneously as a child. She has been conducting research in and among Latin American communities in the U.S. and Latin America for more than 15 years. Although she has conducted research with communities from Venezuela, she is not of Venezuelan origin and has not had the opportunity to spend time in the country. When speaking with participants, she disclosed her personal background and positionality.

In the second case study, the first author interviewed nine participants who are journalists, activists, politicians, and lawyers with ethnic minority identities, and conducted digital ethnographic observations of participants' social media posts on Facebook and X to make sense of posting behavior, social context, and cultural and identity-based complexities present when ethnic minorities in Nigeria are interacting online [55]. The first author's positionality as a researcher from Nigeria informed her deep understanding of the Nigerian socio-political and cultural context, which allowed her to critically analyze themes around ethnicity, repression, violence, displacement, religion, and language that characterize the types of content participants created and perceived to be over moderated or under moderated. The first author is a Black African female from

Nigeria, where she is part of the Egun ethnic minority group, one of the more than 250 ethnic groups in the country [1]. The fieldwork and data collection for this study occurred remotely in 2023 from the U.S.. The study was designed to understand the experiences of ethnic minorities in Nigeria with content moderation when they use social media to document human rights violations and violence against them.

The third case study draws from interviews and digital ethnography examining trans technology creators primarily in the U.S.. Being trans in the U.S. is a high-stakes context, as this group experiences substantial violence, discrimination, harassment, and oppression due to the country's anti-trans political climate [53, 90]. Trans technologies are technologies designed to help address some of trans people's needs and challenges [40]. The third author (a white trans man) collected most of the data, with help from the second author and several other researchers, who were primarily trans people of color. The study was designed to understand trans technology creators' experiences with and processes for creating trans technologies. Our positionalities, and expertise in understanding technology, helped us to build rapport with participants around shared identity and tech-savviness. Here, the second author's analysis is situated in a fluid mix of intersecting field sites, one taking place via interviews where the second author took notes, and the other through autoethnographic reflections based on shared identity and experiences. The question of where the boundaries of a field site in digital ethnography lie depends on the questions that are being addressed and how intersecting dilemmas may exist between the participant and the field site [11, 46]. For this vignette, the second author used textual and thematic analysis of interview transcripts from 104 semi-structured interviews with trans technology creators, 26 of which described the digital practices of transgender social media users and their experiences with platform features and governance. The fieldwork and data collection for this study spanned from 2021 to 2022; we conducted interviews using Zoom, and digital ethnography on social media sites participants used and via participant observation using the technologies that they designed [44, 96]. Finding salient themes in the data related to content moderation, the second author used autoethnographic writing to describe and relate their reflections in reviewing the interview transcripts and field notes alongside quotes from the interviews. The second author is a Vietnamese-American queer and neuro divergent transfemme who wrote the vignette on trans technology creators' case study; her positionality influenced data analysis by allowing her to connect and resonate with the sentiments and experiences shared by trans participants.

4 Findings

In this section, we present three case studies in Venezuela, Nigeria, and North America (primarily the U.S.). Each case analyzes common factors occurring in all three contexts about survival, activism, the lack of a 'voice' and visibility, graphic content about bodies, misinformation but also non-common factors such as misinformation in low-resource languages, context gaps in humor, and geopolitical realities. In each vignette, we use first-person language for analysis and sense-making of the experiences of participants in the respective case studies as well as to narrate the events as they occurred. In the first vignette, we use pseudonyms for participants; in the second and third vignettes, we mostly use participants' real names, according to their stated preference.

4.1 (In)Visible Crisis Work: The Venezuelan Experience

We had scheduled our interview to last 60 minutes tops, but as we reached the end of our time, Jose was still talking, laying out all of the ways he was trying to address the seemingly unending humanitarian crisis in Venezuela. It was the summer of 2018, and I was researching how Venezuelans and Venezuelan Americans in the U.S. were addressing the ongoing humanitarian crisis in Venezuela. Jose, who I met through a Facebook Group dedicated to Venezuelans, was one of the first participants to agree to speak with me. Jose asked if we could meet in person at one of the student centers on his campus. Though an undergraduate student with a full course load, Jose spent much of his time on social media, coordinating support for friends and family back home in Venezuela. During our interview, he apologized for yawning:

> *I get about 3-4 hours of sleep per night. I'm always online, on Facebook, on Twitter, on WhatsApp, I work to coordinate support, raise awareness, post information. But there is so much, and so many different concerns, about what is being posted and who it's going to affect. It takes a lot of time.*

While I had not initially set out to explore content moderation directly, my research led me to examine the efforts of people desperately trying to support their friends and family in Venezuela throughout a prolonged crisis, which naturally intersected with online content issues. Faced with the infrastructural failings of multiple systems they once relied on, participants turned to social media platforms to cope with the challenges they faced. Through my research, I began following several accounts across platforms, starting with Facebook and expanding to Instagram, Twitter, and WhatsApp groups. I discovered that while these platforms facilitated various forms of support, they also complicated experiences of (in)visibility and trauma. In prior work [29], I described the significant time and effort people invested in caring for others at a distance as "infrastructural care," or the collective and often improvised efforts of individuals and communities to compensate for the failures or inadequacies of formal infrastructures, particularly during times of crisis [29].

In the context of the Venezuelan humanitarian crisis, infrastructural care involved using social media platforms and other digital tools to provide support, share resources, and address the gaps left by dysfunctional or absent governmental and institutional systems. This form of care emphasized people's active role in sustaining their communities through both digital and physical means, reflecting a deep commitment to mutual aid in the face of systemic collapse. However, this reliance on social media also introduced new complications, particularly around content moderation and the varying interpretations of what type of content should be allowed online and for whom.

Participants told me that the types of content they posted, which might traditionally be deemed too violent, were necessary to educate the American public about the situation in Venezuela because

official media outlets were not covering what was happening in the country. I met with Renaldo over a Skype video call after finding some of his tweets on a hashtag I was following on Twitter. "*It's like a war is going on in Venezuela and there's nothing on the news [in the U.S.]*," Renaldo threw his hands in the air in frustration. "*It's like CNN's talking about, like, you know, who had sex with who, you know, while this is going on.*"

Frustrated by his American friends' lack of knowledge and sometimes disbelief over what was occurring in Venezuela, Renaldo decided to start posting on Facebook. He purposefully chose posts with images or videos to "prove" to his American friends that this violent conflict was actually happening. While his American friends thanked him for informing them, Renaldo's Venezuelan family was offended and found the content inappropriate and shocking. His uncle, for example, asked Renaldo to stop posting that content because he didn't want Renaldo's teenage cousin, who was also on Facebook, to see the violence that was occurring in their home country. Renaldo said his posts were occasionally flagged or removed, which he guessed was due to Venezuelan family members or friends in the U.S. reporting content that they did not want to see. Renaldo viewed this less as a platform issue and more as an issue of context collapse, since he had multiple audiences as Facebook friends.

One way Renaldo managed this tension was to publish a post for 1-2 days and then remove it, so the post did not remain on his Facebook page permanently. He said that he felt he had to post it on Facebook instead of another platform because Twitter did not reach his personal audience and, by making the story personal (i.e., sending it out as himself to his personal audiences), his friends, particularly his American friends, would pay more attention and believe it more. Renaldo said that he felt he had to do this work himself, because no one was going to speak out for Venezuela. Simultaneously, Facebook's design contributed to making his labor more complicated, leading to relational consequences. While Renaldo struggled to balance conflicting reactions to his posts, other participants struggled with their own exposure to violent content, again developing personal strategies to manage their online efforts.

Mercedes, for example, developed tactics with her spouse to help them manage exposure to violent content while also staying connected to online support efforts. Although she tried to avoid disturbing news stories on Twitter and Instagram, if Mercedes wants to maintain contact with people in Venezuela, she cannot disconnect completely, especially from WhatsApp, where she is a member of several private groups with hundreds of members. If she wanted to maintain her role in supporting people in Venezuela and feel like she was involved, she has to be on WhatsApp. Since Mercedes had to move within the same platform in her efforts to confront the crisis, she and her spouse warned each other if they come across a violent image or disturbing news in these groups. They tried to comment back to offer support while also avoiding exposure to all the heartbreaking stories. If one of them saw a violent image that they knew would upset the other, they reacted to the message with a rare, preselected emoji. If the other came across that message while scrolling through the groups and saw the emoji, they knew it had been tagged and not to open it. Mercedes explained:

> *We had more homicides in Venezuela than any war. At that point we even had more dead people than the Syrian war. So that kind of thing and of course anything that has to do with torture…I avoid that part. I'm like an ostrich; I mean, I put my head in a hole. I don't want to see that, I don't want to know anything about that…Because, there's nothing I can do. It's going to make me feel bad. It's not going to help me in any way and it's just the reality that I already know it exists and that's the reason why I left.*

This sentiment was echoed by several participants, including some who said they have to shut off their phones for several hours, or even a whole day, because they become too overwhelmed by the violent content. Some participants described the violent content as necessary to the cause of educating and forcing others to bear witness to what was occurring in Venezuela. Further, some participants said that sharing violent or troublesome content also served to show their friends and family back in Venezuela that they had not forgotten them and were in solidarity with them. People in Venezuela were grateful for others who were speaking out, but many Venezuelans who had already left the country viewed this content as re-traumatizing.

For many participants, too, as more time went on, they described feeling burnt out from engaging with *"the most extreme stuff,"* as Jose described it. However, they did not want to block people entirely because they understood, in part, that posting this graphic content was necessary to raise awareness and, mainly, because the people posting were their family members and friends.

Interestingly, the content that is typically considered "safe" on Instagram and Facebook, especially from a privileged Western perspective, was deemed problematic and even dangerous for participants to post. For example, Renaldo said that he no longer posts pictures on Facebook about his daily life and instead reserves that for his small group of followers in Venezuela. According to Renaldo, *"I don't necessarily want people in Venezuela knowing that I live [in the U.S.] or that sometimes I go on vacation and travel because it creates a sort of sense of, you know, negative sort of envy I think, but also… I don't want to showboat on those people who are hungry."*

Several participants voiced concerns about platforms amplifying the "wrong" content, such as making photos of vacations and displays of wealth overly visible, while suppressing posts from people in need. Even more critically, however, if participants were perceived as having money via the identity they portrayed on social media platforms, their family in Venezuela could become a target for kidnapping. Due to the extreme poverty in Venezuela, desperate groups started scouring social media sites, trying to locate "wealthy" family members outside of Venezuela who could meet ransom demands. This led to participants, like Valentina, restricting the type of content they shared because *"no one is going to look out for people in Venezuela online, so I have to be the one who is careful with what I post."*

In addition to self-censoring the content they shared to protect family members and friends, participants also voiced concerns and confusion related to misinformation. Certain content was flagged and removed as either inappropriate or untrue. Still, conflicting news on multiple platforms led to the removal or suppression of

posts from people attempting to warn others about potential dangers in the country. The lack of context-appropriate oversight over information accuracy posed significant challenges, especially after the Venezuelan government closed independent news organizations and began using social media to broadcast its position. I began following Venezuelan President Nicolas Maduro on Instagram and Twitter, as well as related hashtags, to monitor the type of information the Venezuelan government was disseminating. I encountered a flood of contradictory content from official government accounts, Venezuelan citizens in the country, and members of the Venezuelan diaspora. The Venezuelan government flooded platforms with content, overwhelming moderation tools in an information war. Even now, I regularly receive notifications on Instagram saying, *"Nicolas Maduro is live"*, as the president frequently uses these platforms to convey his messages.

Unable to trust the information released by their government, participants turned to social media to verify the accuracy of news. During my interview with Jose, he described how people used social media to try to verify and spread information to protect loved ones; however, this had unintended consequences:

> *Social media became important during student protests. Soldiers went after students, so the students started to take shelter… in the homes of civilians. So the soldiers said, 'okay, now let's attack the civilians so we can stop the students.' So, information infiltrated that they are going to come to this building or that neighborhood. So, the people started to use social media in one way or another, and that created panic among the people that were in Venezuela and abroad, as well, because in Venezuela, there are many controls as far as Twitter and Facebook and they tend to block many, many web pages that are totally available [in the U.S.].*

This reliance on social media, however, also concentrated power within these platforms and, by extension, in the hands of governments. When a society's primary source of information becomes a few centralized digital spaces, the ability to "flip a switch" and block access to these platforms becomes a tool for control. As a result, individuals found themselves in a precarious position where pursuing information could lead to further panic. Jose said it became unclear whether his family in Venezuela desired the information or not, as it could incite greater fear in Venezuela. The fear of government retaliation and the psychological impact of disseminating potentially fear-inducing information became an additional burden for those attempting to navigate the information landscape. The dangers of perception and fear of government retaliation were ever-present. During Skype conversations, participants expressed concerns about kidnapping and being careful about what they posted online. As one participant, Natalia, noted:

> *And so, yeah, I think there's this fear, like underlying fear, like how can the government use any information that you post on social media, because there have been many, many stories of people that are violated in that way. Like I think it could happen in any country. If you go to the Middle East– like I'm not going to be posting, "I'm traveling to the Middle East." You know what I mean? It's being aware of what you're sending out to the public, is definitely in the back of my mind, at any time.*

Participants faced issues of content removal and self-censorship, as regular people often had their posts removed while trying to navigate unclear policies across multiple platforms. Jose, for example, wrote blog posts on Facebook to educate others about how the current government was treating the Venezuelan people. In some of the posts, he shared pictures to try to lend credibility to his argument; however, posts like this were removed for violating Facebook's terms of service. Jose guessed that it was either someone in the Venezuelan government who had reported it, "or", he added, *"it might even be a family member of mine. My mom is scared for my safety and doesn't want me to speak out online anymore."*

The confusion described by Jose was common across participants as overlapping audiences and motivations collided in online spaces. Some believed it was because people were offended and reported them, while others thought the Venezuelan government was trying to silence dissent. However, Jose, like other participants, felt that it was up to them to combat misinformation and inform the public as to the situation. This created additional labor and precarity as they sought to share information. Similar to strategies described in related work [61], participants developed hashtags to circumvent censorship and prevent their posts from being removed. However, concerns about over-exposure and the labor involved in these efforts persisted.

The complexity of this situation lies in the fact that platform needs are not only culturally specific but also individual and personal. Participants had to develop their own networks and infrastructures, turning to social media as a form of critical infrastructure to support information and resource needs. However, they soon realized that this approach was unsustainable due to the substantial labor involved. Several participants began to formalize their tactics to create more sustainable methods of support and emotional coping as they faced overlapping power structures on and off platforms, like Mercedes and her spouse's collaborative coded mechanisms described earlier.

These issues are not confined to Venezuela or the U.S. Although this study began in 2018, the relevance of these concerns persists and, in many cases, has worsened. For example, in the case of Venezuela, people posted information online to try to warn those in Venezuela about violent outbreaks. If that content had been removed because it was flagged as violent or as misinformation without being thoroughly vetted, this information may not have been disseminated in time to save people's lives. This raises critical questions about platforms' responsibilities and the demands placed on content moderators when dealing with crucial issues that require specific contextual and cultural literacy. It also exposes the additional labor placed on individuals who are already in precarious positions.

4.2 In Nigeria, Social Media Use is for Life's Sake

The death of thousands of people in his community leaves him no choice; for Reuben, a community leader in the Southern Kaduna area of Central Nigeria, his social media accounts serve two purposes: to post jokes, and to post images of severed limbs, decapitated children, and people in his community after they are attacked by

Fulani herdsmen or state actors. With a detached ease that concealed the psychological and emotional burden of his work, he tells me which social media platforms are best for posting dead bodies: "*It's more tolerant where you wanna [show] pictures of dead bodies, you know, of mass burials and some other things. I go to Twitter (X) and I post those ones there*".

Reubens's folk theory of the best social media to accommodate the gory images he is posting stems largely from his lack of access to mainstream media as an ethnic minority in Nigeria and his experiences with content removals. This is consistent with research that shows that social media users develop folk theories to help them make sense of their experiences on social media [66].

The outlet that social media provides for him to document his experiences with human rights violations and violence targeted at ethnic minority groups is multipurpose in its ability to bypass state media censorship against minoritized populations to reach broader audiences and to bear witness to the violence he experiences in real-time:

> *I use my social media to vent my anger, to express myself, you know. Social media gives me an opportunity to speak freely [and circumvent] the censorship, the restrictions that [we] usually have in the mainstream media. And of course, it's cost-effective to [use and] gives me direct access to some of the target audience[s] that I want to reach.*

This type of access is crucial for Reuben, who works with rural communities in Southern Kaduna where people have limited access to digital literacy, electricity, and digital infrastructure.

As we are speaking on the Zoom call, Reuben's electricity goes off and it becomes dark in the room where he is sitting, infusing a sense of urgency into his voice. He wants to know if I can still can still hear him. I can but it is hard to see him. I make out his figure in the dark room as he struggles to make himself visible to me. In many ways, that moment metaphorically reflects his reality as an ethnic minority using social media. He goes on to share an example of a time when he believed that he experienced content and account removal on Facebook without consideration for his unique identity as an ethnic minority in a Global South country trying to share crucial content related to human rights violations. On his way back from an Internally Displaced Persons (IDP) Camp, a place where people especially ethnic minorities displaced from attacks go to wait for medical help, aid and seek temporary refuge, Reuben encountered two victims of ethnically motivated violence, which he decided to take a picture of and document on Facebook: *"...I was trying to [show that] this is the plight of some of these people in a camp, you know?"* Shortly after he made the post depicting the conditions of children in the IDP camp, Reuben said Facebook suspended him for showing nudity. Internally Displaced Persons Camps in Nigeria are typically filled with victims of ethnic and religious violence who have freshly fled their villages and lost everything, including clothes and shoes, awaiting aid from emergency services or aid groups. While Reuben says his aim was to amplify the violent and desperate conditions his community faces, Facebook's moderation algorithms, triggered by the photo of children who were probably scantily clad, assessed the photo without context and quickly suspended his account.

Reuben's experiences also point to Western social media platforms' lack of knowledge and consideration of non-Western cultural contexts, which leads them to impose Western norms and ideals via their moderation policies in geographic and cultural contexts where such norms do not apply. In another example that shows the inadequacy of such one-size-fit-all moderation approaches that conform to Western ideals, one of Reuben's posts on Facebook was moderated for making a self-deprecating joke about Nigerians: *"I said how we in Nigeria are crazy, you know, that was what I said"*. To Facebook, though, this post was taken down for *"racial profiling"*. *"Oh, wow, what I said was more like a joke, you know, it was more like a joke"* he told me, shocked and confused that his joke which would have been understood by any Nigerian was taken out of context. Because Nigeria is a largely racially homogeneous country where ethnicity and religion, not race, are the drivers of socio-political divisions and discrimination. A post by a Nigerian making a joke about Nigerians being labeled as racial profiling shows just how large the context gap is between Facebook's content moderation policy and the Nigerian cultural experience.

Alheri, an ethnic minority from Southern Kaduna and a community organizer, is best known for her work documenting displaced victims of ethnic violence in her community. I met her a few years ago when I was a journalist and was reporting on the killings in her community. On Facebook, she has a following of about 14 thousand people; a scroll through her page reveals a mix of personal content often interrupted by documentation of violence and pleas for aid for displaced members of her community. For Alheri and the people she is a voice for, social media is a tool for survival, without it, they will be no way to document their dead, or sustain those who have survived the attacks. Her work's importance is evident when she highlights the vulnerability of the population she is documenting:

> *As of 2019, the official count [of Nigerian people in IDPs] was 12,480.... I was amazed at the number, and we hadn't even finished counting. Remember, this is [just] one local government in Southern Kaduna. So I was alarmed, why are they not being helped? [The villagers] have no phones. They're not on social media. Nobody knows what they're going through. So we get pictures and then we post [about] what happened, on so and so date. There are a number of vulnerable kids who need medical attention, and one thing we advocated for, and [are] still advocating for, was for, at the very least, for the government to provide a hospital [so] that when [there is an] attack, they can go there for treatment. Because I mean, they attack you, and then everything you have is gone.... Normally [as a] human being your neighbor helps you. But [when] the whole village is attacked, the whole village is down. All the money you saved is burnt or stolen. All your food is burnt or stolen. So who helps them? We've had cases where people died or limbs had to be cut off because they don't have less than a dollar [the likely cost at a public hospital] to treat their wounds.*

Despite the important nature of her work, Alheri frequently gets content violation notices from Facebook, which she perceives as arising from a disconnect between Facebook's moderation policies

and the unique nature of her content. In an experience that is very similar to Reuben's, Alheri described how she posted an image of someone who was in a dire condition in the hospital after an attack. In the picture, he had only boxers on. Facebook sent a notification that she had violated the platform's nudity policy:

> *They [Facebook] made it look like it was a nude picture, but it wasn't a nude picture. It was someone that was sick…It was definitely someone that was ill and had an injury, but then it was interpreted as [though] I [had] posted a nude picture. So I was given a warning. From then on, I would always use [a] smiley emoji to block out an injury.*

Such experiences altered Alheri's posting behavior and made her much more cautious and "proactive" – she had to carry out additional labor to avoid moderation. The lack of a tailored moderation system and policies for people like Alheri and her community directly impacts her ability to use social media for her community's distinct needs for self-preservation and visibility.

Makut, an ethnic minority who is also a local politician and community organizer, told me that he made a post on X after a Senator who is an ethnic minority was killed while attending the mass burial of members of his constituency who were killed in an overnight attack on their village – by the same attackers. Under his post, people left comments in the Hausa language saying, *"It's a good thing that happened, yeah, I'm happy this happened, let them be killed more"*. He believes *"…that's just targeted hate"* that X refused to moderate. In his folk theory of why such inciting comments under his post remained online, Makut said he believes the perpetrators of this online violence use the Hausa language, rather than English because they know it is a way to bypass X's moderation. The lack of cultural and linguistic competency and capacity by social media companies like X to moderate posts in fairly popular languages like Hausa which is spoken by about 88 million people in West Africa [2] emboldens bad-faith actors to perpetuate online harassment and heighten tensions in moments of ethnic or religious conflict. Makut's frustration with how hate speech in Nigerian languages are moderated is compounded by seeing that social media companies have translation tools integrated into their apps, yet fail to use these same tools to moderate content in low-resource languages: *"I am always wondering, since Twitter allows you to translate, [and] also Meta [has tools] that allow you to translate. So it can also help them to find hateful comments, and I'm wondering why they were allowed such posts to still be hanging online"* .

Content moderation that is equitable and rooted in context – which could substantially help minoritized communities in violent situations – cannot occur when users and social media companies are literally and figuratively speaking different languages.

4.3 Bodies and Genders Under Attack: The Trans Social Media Experience in the United States

When I was a teen on the Internet years ago, my gender always came into question. *"Is the Hibby a boy or a girl?"* users in a Pokémon Online chat would ask each other, and I would never answer. A comic made by one of the mods showed a figure meant to be me with a question mark on their chest in front of two bathrooms, marked with male and female gender signifiers. I was around thirteen at the time, and no one vocalized anything about how these commentaries on my gender were inappropriate. Transness was not at the forefront of the community moderators' minds, and even though I was a mod myself, a thirteen-year-old who had not even discovered her own transness knew little about the subject either. This speaks to the importance of trans contexts in moderation [96]. Now, around fifteen years later, I have the insight and understanding to know that a need to identify and label me as one category within a gender binary or as inherently 'other' was at the forefront of these events. I also have more experience with gender transition, and still feel this push-and-pull offline, as others look at my feminine gender expression with makeup and clothing, alongside traditionally masculine aspects of my appearance, such as my mustache and Adam's Apple, and radiate confusion. While the questioning does not happen in the physical world as much as it did online, I am still frequently forced into one category or the other in a gender binary. Quotes from our interviews with trans tech creators resonate with many of my own experiences as I read through the interview transcripts. Many of these words are familiar, as I have heard similar stories from my friends or from other trans people on the Internet. This resonance and familiarity speaks to sort of trans universality in interviewees' experiences, and to my strengths in writing from an in-group position that encompasses both closeness and my own unique experience.

Trans experiences in online spaces are heavy with both over moderation and harassment, which makes content moderation especially salient in considering the intersections of trans oppression and technology. Though trans people, unlike our other two case studies, are a diverse group of people dispersed across the world, for those of us in anti-trans political environments like the U.S., it is a high-stakes, crisis moment. Trans people in the U.S. are facing unprecedented levels of hate, harassment, harmful legislation, violence, and marginalization [53]. This anti-trans oppression extends beyond just the physical world, and permeates online spaces as well, perpetuated by both everyday users and by those in power. For instance, in January 2025 Meta CEO Mark Zuckerberg announced that the company would "remove restrictions on topics like immigration and gender that are out of touch with mainstream discourse," citing "recent elections" as the rationale [71]. One consequence of these lifted restrictions is that users across Facebook, Threads, and Instagram may now use derogatory language toward LGBTQIA+ people.

"As trans people, we're constantly told that our perspectives aren't valued…" Sophie Debs, the creator of browser extension Gender Neutralize (which "de-genders" online language) told the interviewer. *"It takes a lot of bravery to voice your opinions when no one says it matters"*. I think about the thirteen-year-old Hibby and wonder what would have happened if she told others to stop discussing her gender. Would anyone else have supported her, or would she have been alone in confronting transphobia? Too often online, users do not have the power or voice to speak up for the injustices they face, or to appeal incorrect or unjust content moderation decisions [69, 97]. *"I think with minority populations… the most effective way for us [to be heard] is people who belong to that group and are on the server vouching for it"*, Laur Bereznai of online community Trans Peer Network told their interviewer. If the thirteen-year-old Hibby

spoke up, she likely would not have had support, because unlike the Trans Peer Network, the space she was in lacked other trans people who might chime in to support her and hopefully change the space's moderation policies.

But it is all just online, right? Thirteen year-old second author Hibby is safe in the physical world, surely?

These three trans influencers were attacked on Hollywood Boulevard, and that's known to be a liberal era… There's literally signage on the street that said, "Trans lives matter", yet these trans influencers were attacked, and no one helped them… There's nothing in place for the safety of trans people in public.

Nikki Nguyen of the Trans Defense Fund says this to her interviewer and I think about the man in Chicago's Union Station who called me a faggot chink and told me he would kill me. Although not openly trans in my expression at that time, this man still saw something offensive in me. Moments like this are not surprising for trans people, who face high rates of violence in the US [53]. When your life is compromised in physical spaces, should online spaces serve as a reprieve? It would be ideal if so, but trans people are among the list of marginalized groups who face disproportionate content moderation on platforms [42], as well as increased hate and harassment online [84]. So then, how do trans people survive? How do we implement content moderation systems to reduce harm in these contexts of survival?

"Trans people have been using survival strategies for centuries, that look like technologies that we use today, and doing things like rituals that look very much like algorithms" says Micha Cárdenas, author of Poetic Operations [14], to her interviewer. These survival strategies Cárdenas mentions remind me of how I would dress more traditionally masculine to look less trans to people, or how I would find small enclaves of queer and trans people and people of color to truly express myself. It also speaks to how trans people online have found ways to work around algorithms or, sometimes, to be less visibly trans online [5, 23, 65]. However, platform features can also render such visibility harmfully, as seen with Facebook's "real name" policy—which sparked controversy among trans users who could not legally change their names to meet the platform's definition of "real" or "authentic," and which also disproportionately impacted other groups such as Native Americans and abuse survivors [43]. What is considered "real" or "authentic" continues to shift even today, with the executive order that the U.S. only recognizes two sexes, male and female [47], as well as the marking of trans citizens as their assigned sex at birth through passport gender marker changes following this executive order [15]. As seen with all of these shifts around what is "real" or "authentic," platform policies reproduce a colonial, state-centered approach to identity that continues to proliferate in physical-world contexts as well. *"We as a community don't have the resources we need in order to have more technology that is actually for us. Usually, if we want to get money for them, we need to get money from cis people.… It makes the whole thing more complicated"*, says Bereznai of the Trans Peer Network. A lack of resources by and for trans people means that trans people have to create their own resources and survival strategies, or risk being co-opted by non-trans developers or stakeholders, who may not fully understand trans contexts.

My mind shifts to when I created a queer gaming space on my undergraduate campus, and the premiere gaming space there questioned why we had created that space, saying it was unnecessary. I later saw their space in action when I joined their Discord server and saw various homophobic and transphobic slurs used by people who were not queer or trans themselves. Here, yet again, we see a lack of voice for a marginalized group, whose concerns are considered as superfluous or unnecessary. Trans contexts are those that revolve around trans people's online safety and well-being, which cannot be fully served by technology, but could at least be served partially or better than they are now by one-size-fits-all approaches to content moderation and platform governance [84, 96].

Trans people face problems in content moderation because of hate and harassment, but also because of how trans bodies themselves are policed [65]. *"Even if I post images of underwear, I think their [Facebook's] bots just pull the stuff down. Anything that vaguely looks like underwear or a penis or anything like that, without a human ever looking at it, their bots basically pull it"*, says Leo of trans prosthetics company TransFormaGear about Facebook's moderation of TransFormaGears' posting about their merchandise. Trans prosthetics and undergarments are an important part of many trans people's lives [76], but one-size-fits-all approaches such as those employed by Facebook/Meta lack the context to understand that a prosthetic penis is an important tool for trans affirmation, not obscene content in the traditional sense. These issues related to the body and supposedly "graphic content" show how mainstream social media platforms can invalidate and erase trans bodies and access to important tools for transition.

How can trans people learn about their own bodies, if other trans bodies always remain hidden out of view? Avery Dame-Griff, the creator of the Transgender Usenet Archive (an online archive of posts from early trans online communities/newsgroups), discussed how the presence of adult content in the archive limits where he can house the archive. Yet he acknowledged how important such material was: *"It's got adult material… but it should be respected as important."*. Here, Dame-Griff is speaking on how erotic trans media can sometimes be influential and educational for trans viewers [41], but how some online repositories may balk at such material being included in an archive. I think about starting my hormones and the lack of knowledge I had about my own body during its changes, and the lack of access I had to erotic trans media that did not outright call me or the trans adult actors "tranny" in the title.

Supporting trans contexts in content moderation requires acknowledging the survival methods and lack of voice trans users have online. Further, platforms must acknowledge trans users' unique contexts, such as the physical world harms they face, the need to share information and media on trans bodies for education and awareness, and the importance of trans-specific spaces.

5 Discussion

The three high-stakes contexts we study highlight the experiences of marginalized populations who are using social media in unique ways that their identities and circumstances require, yet that do not align well with common and expected social media uses. Participants in each of these case studies use social media to construct alternative infrastructural realities in response to a breakdown of

systems that typically enable visibility, such as mainstream media and traditional physical-world archives. We presented examples of breakdowns in physical infrastructures of visibility, documentation, and community building, which led participants to adopt social media practices that operate on the peripheries of normative or socially acceptable social media use. We argue that the platform features participants used intersected negatively with their identities and circumstances, leading to the de-prioritization of their infrastructural and contextual needs. We further argue that the inherent lack of complexity in social media platform design, particularly in accommodating high-stakes use, is rooted in design decisions that mimic physical world power asymmetries, leaving marginalized populations as an afterthought. We highlight how this design gap leads marginalized populations to perform labor and bargain for visibility in ways that further entrench their marginality.

Prior research highlights how geographically-separated people experiencing humanitarian crises use social media to create infrastructures of care [29] and how marginalized populations use social media for archival purposes in high-stakes contexts [55]. This paper uncovers how social media platforms designed on Western hegemonic practices collapse when marginality, identity, and high-stakes contexts intersect. In this Discussion, we draw from scholarship on repair work to speculate what repair might look like for marginalized populations using social media in high-stakes contexts. The work of repair extends and safeguards capabilities that are prone to disrepair and decay [51]. Repair work is constructed in time and space, occupying intersections where socio-technical systems show instability and an overall lack of structural integrity [51]. Further, we advocate for integrating, incentivizing, and mainstreaming the marginalized populations' alternative structural design realities into social media platforms, while also recognizing that social media companies have little incentive to do so and continue to neglect and take advantage of the most vulnerable populations.

In the next sections, we highlight examples of instances when platforms broke down for the marginalized users in the cases and reveal places were their practices present opportunities for repair, articulate how platform breakdown make social users on the margins vulnerable stakeholders and highlight how participants in different contexts had diverse repair needs.

5.1 Repair Work In High-Stakes Contexts

In the beginning of this paper, we likened platforms to buildings: social media platforms with design features that do not accommodate the peripheral use of social media in high-stakes contexts where identity, labor, and marginality intersect are like buildings in need of repair work. Participants in all three vignettes scaffold their lived experiences and identity onto these platforms in unconventional ways that challenge existing design. Prior research shows that while the design of systems that facilitate automatic content moderation can sometimes be beneficial for sanitizing offensive content in digital spaces, these same systems perform poorly when navigating tensions around what is deemed contestable or offensive, in part because those judgments reflect the norms and biases of the human raters who supply their training data [9]. Systems in need of repair create opportunities to explore overlooked sites of innovation in ways that support re-purposing existing infrastructure to benefit new communities and people [52]. We adopt Jackson et al's [52] view that the breakdown of sociotechnical systems provides an opportunity for repair, and extend it by showing this in a new breakdown context with new opportunities for remediation: marginalized populations using social media in moments of heightened precarity.

Consider Renaldo's use of social media to bring visibility to the humanitarian crisis in Venezuela. His disenchantment with a U.S. news echo chamber that does not prioritize reporting on Global news affairs led him to create Facebook posts. While his posts made the conflict in Venezuela visible to Americans in his community, it was offensive to others in his community who complained about it. This led to Renaldo balancing trade-offs between the relational cost of offending family members and the need to raise awareness about Venezuela. It also created tension within his family, since Renaldo suspected some of his posts had been flagged for removal by his family members. Rendaldo's case reveals a flaw in the design of the report feature that helps to manage platform governance. Because platforms designed around governance do not necessarily account for a multiplicity of contexts, Renaldo's work is fraught with navigating this failure in design, adding to the labor he has to perform to be online, given his commitment to sharing about the crisis in his home country. The report feature failure further reveals that in high-stakes contexts, flagging content transcends reporting offense and enters into the realm of care. For example, Jose and other participants in the Venezuelan case study described the feeling of burnout from constant engagement with sharing and viewing extreme content. They understood that they must *bend* tools like the report button or the block button to suit their infrastructural needs. Renaldo understood why his family would flag his posts for violent content, because of how that content might impact his teenage cousin. These flagging behaviors were neither inherently good or bad. His approach to repairing this breakdown by deciding to publish a post for 1-2 days before removing it was pragmatic and necessary. We highlight how workarounds like unexpected uses of the report button can be instructive for platforms seeking opportunities for repair in the design of platform features.

In the Nigerian context, the failure of platforms like X to moderate content due to a lack of capacity for low-resource languages in the Global South manifests as heightened hate in moments of crisis. Makut, who is an ethnic minority, relies on social media to document things like attacks on his village. He has to deal with hate comments left in the Hausa language because, according to his folk theory, perpetrators of targeted harassment know that they can bypass moderation because they are posting in languages that social media companies like X do not have the capacity to moderate. The current approaches to address power imbalances in content moderation policy in the Global South further deepen mistrust because users perceive human moderators (proposed as alternatives to automated moderation) often have little to no competency in local cultural norms or languages [86, 87]. The ways platforms ignore languages and contexts they deem negligible shrinks our ability to assess harm to marginalized populations when majority groups' experiences largely define harms. It leaves little room for the complex experiences of peripheral social media users.

In the U.S. context, Leo, the founder of the trans prosthetics company TransFormaGear described how content of images such as underwear or products resembling genitals gets taken down on Facebook. This blunt moderation constitutes a form of erasure for Leo and other trans people who rely on prosthetics and undergarments for trans affirmation. Yet because the platform governance features consider the type of content that Leo posts as sensitive, the content is removed. Research into the experiences of transgender people with content moderation show that even when abiding by platform guidelines, disproportionate content removals persist [42]. Trans social media users are more likely to have their content classified as "adult" content, which resonates with Leo's experience [42]. His content and those in the remaining two case contexts falls within what Haimson et al. [42] describe as moderation gray areas: content that are technically allowed according to platform rules, but transform into gray areas during content moderation enforcement because of a lack of contextual information [42]. Marginalized groups in precarious conditions have a complex understanding of gray area content that is often imperceptible for those moderating content. Learning from marginalized groups' understandings of these moderation gray areas is another way to transform marginalized groups' social media practices into repair work.

Lucy Suchman indicates that the boundary between machines and humans is determined by the interface and the differences the interface is meant to bridge [93]. However, when the interface fails to bridge those differences, people must step in. Jackson argues that when systems inevitably break, the subtle acts of care and maintenance that restore order constitute repair, a core, generative, and highly skilled form of labor often rendered invisible [52]. The labor performed by our participants—from developing community governance to using "algospeak" to circumvent censorship—is not merely reactive coping. Instead, it should be recognized as a sophisticated act of sociotechnical repair that sustains communal bonds and critical resource management in the face of platform failure. Their work is a non-normative yet essential form of innovation that bridges the gap between the platform's stated goals and the precarious realities of their lives. By performing this infrastructural care [29], they are continually reconfiguring and reassembling the platform's functionality to meet urgent, life-sustaining needs that the corporate design failed to accommodate [51]. In the foregoing paragraphs and our metaphor, we discussed how the building that is platform design is badly in need of repair work. Next, we highlight that complexity in platform design arises because of the occupants of this building in need of repair have distinct needs and motivations.

5.2 The Precarious Stakeholder

Stakeholders who influence how a platform is designed and deployed in high-stakes contexts include platform owners, site designers, and employees. Marginalized platform users are also important stakeholders, yet they do not have input or influence into platform governance and design. Repair work extends the capabilities of a system to take on and work to mend the consequences of decay [51]. The work of repair inherits old sociotechnical systems and all the ways that they have failed to function at the margins and at intersections to make new systems that work better [51]. Yet, competing motivations mean that repair work is fraught with complex considerations and uneven power dynamics that may hinder repair work.

Platform motivations that do not sufficiently take into account the experiences of precarious users when designing and deploying technology in high-stakes contexts increase the labor burden that these marginalized groups must undertake. In the Nigerian context, Alheri uses social media as an infrastructure of care to crowdfund, create archival records, and seek help. Yet the lack of tailored moderation is a form of decay that further plunges her into precarity. When the content she posts is taken down because it does not fit neatly into the rules that content must abide by, she cannot help her community or create online records of her experiences. Her pro active action to self-censor posts in order avoid moderation adds to the labor she has to perform remain visible online. Modifying her posting behavior serves to repair the type of one-size-fits-all moderation policy that led to the content violation policy she received. Marginalized social media users subjected to platform policy changes that target their identity, like in the U.S. trans context, reflect how expediency is built into platform design. Policy and design practices that render the rights of marginalized social media users dispensable in light of social, political, or economic considerations are colonial in nature and replicate real-world asymmetries. While platforms sometimes facilitated forms of support and visibility for people in high-stakes contexts, they also sometimes caused invisibility and trauma. For participants, the constant negotiation and deliberation of what to post to manage internal and external tensions and consequences increases the labor they must perform. We argue that this way platforms impose this type of labor is violent, as it substantially negatively impacts marginalized populations and limits their ability to provide much-needed support for their communities. Research already suggests that marginalized populations must perform invisible digital labor to circumvent economic precarity [56]. We extend this by showing how the labor that marginalized populations in high-stakes contexts must perform reinforces a type of neocolonial violence that further entrenches their dispossession from systems of communication and records.

In all three of our cases, participants relied on social media not just for personal expression or low-stakes browsing, but as a form of critical infrastructure - as a way to coordinate survival strategies by amplifying their, and their communities', voices and engaging in activism and advocacy. Yet, in all three cases, their platform use was constrained by rigid governance structures that failed to recognize the urgency and necessity of their content. Whether navigating censorship, misinformation policies, or algorithmic erasure, participants found themselves forced to work around systems that were not designed for their contexts. Their content was in some ways illegible to platforms – often because it included, by necessity, "graphic" or sensitive content documenting violence, human rights violations, or lived experiences of marginalization. This disconnect reveals a deeper issue: platform governance is not simply about content moderation, and platform design cannot be done in abstraction. The design of platforms and platform policies determine who gets to be seen, heard, and supported in moments of crisis and political upheaval. We posit that social media users in high-stakes contexts reveal hidden truths about the (lack of) structural integrity

of governance systems and structures. The malfunction that participants experienced is emblematic of systems that fail to perform as intended when high-stakes contexts arise. As Shahid and Vashistha [87] argue, this constitutes a form of digital colonialism, where a monolithic moderation infrastructure, heavily influenced by the Western notions of free speech and social norms, is imposed globally. This imposition disregards the sociopolitical and contextual realities of users, effectively reinforcing global power asymmetries and allowing the platforms to insert themselves into local conflicts in ways that often compound racialized and social-political oppression [87]. The complex, high-stakes communication needs of our participants, whether documenting crises or self-expressing against binary categories, are systemically deemed peripheral to the platform's core, Western-defined operational model, thereby deepening marginality instead of mitigating it.

5.3 One Size Does Not Fit All

Despite similarities in survival, activism, visibility, and graphic content about bodies, each context is unique. For instance, the Nigerian context requires attention to rurality, literacy, access to digital infrastructure, and the remaining traces of colonialism. In the Venezuelan context, primary issues involve misinformation, the need to provide visual evidence proving that violence is occurring, and government overreach in flagging users' political content for removal. And in the U.S. trans context, critical tensions include the overwhelming need for online safe spaces but the difficulty of properly vetting them, the overblocking of images of trans bodies, and the untenable combination of physical world violence and online harassment. These differences make clear that it is not enough for platforms to create new policies, algorithms, and traditional trust and safety staff positions dedicated to moderation for "marginalized" users and contexts writ large. For instance, a content moderator with expertise in trans topics would likely have no knowledge on how to interpret online harassment written in the Hausa language in Nigeria. Similarly, someone well-versed in the Hausa context could not be expected to make correct decisions when encountering potential misinformation surrounding Venezuelan politics. In the next section, we recommend design considerations that can be applied universally but can be customized to especially account for the needs of marginalized populations in high-stakes contexts.

5.4 Implications for Design

I think with minority populations. . . the most effective way for us [to be heard] is people who belong to that group and are on the server vouching for it. Laur Bereznai, Trans Peer Network

We recommend that platforms integrate alternative infrastructural designs that users in high-stakes contexts have developed in response to perceived platform failures. For example, AEGIS, a tool created by Rosa Chapperri, allows a network of LGBTQIA+ Twitch content creators to collaboratively vet users and alert others of transphobic accounts. We recommend that platforms collaborate with marginalized users to integrate alternative infrastructural features and systems *that these groups are sometimes already creating*, such as through workarounds and creative sociotechnical repair work.

5.4.1 Emergency response tools for marginalized social media users. We recommend that social media platforms build contingency features and tools into their platform and moderation design to allow people experiencing heightened precarity to construct experiences that makes the most sense for them depending on their context at any given time. This tool can track the needs of people and communities experiencing precarity in real time, ensuring that resources are allocated to help users safely use social media in moments of upheaval. This type of tool building is not new. For example, Facebook's Emergency Disaster Response, which launched in 2014, is intended for users to conduct a Safety Check that helps them stay in touch with family, friends, and community in the event of a natural disaster [58]. We propose extending this type of design for content moderation decisions and platform features. Specifically, we recommend that an emergency response tool be built into social media platforms. This tool could serve as a dispatch, allowing automatic designation for users in high-stakes contexts and allowing users to opt in based on their unique needs.

5.4.2 Exit mechanisms for wellbeing in high stakes contexts. We further recommend that social media platforms develop and integrate tools that support personal and community well-being during high-stakes events. Participants like Mercedes and Renaldo described employing tactics to help them manage exposure to the extreme content that they were posting; a mix of communicating through pre-selected emojis to communicate distress, posting and deleting, and understanding when their content is reported by members of their community. The ways participants appropriate platform features to support their mental and emotional well-being indicate the potential for a dedicated feature or tool to the type of support they seek. We recommend that social media platforms provide users with exit buttons that allow people in places where high-stakes events are occurring to easily opt out of viewing graphic content, and to just as easily opt back in as their conditions and feelings change.

Our recommendations extend prior research on value and repair, which shows that repair can take on modes of human interaction with technology as a continuous process, never fully realized [49]. What social media users deem valuable in platform design becomes apparent when they attempt to use social media in high-stakes contexts. Social media platforms often attempt repair work after harm has occurred – retroactively deploying resources, apologizing, changing the design of platform features – to the detriment of marginalized populations. Instead, we advocate for platforms to preemptively adopt and incentivize alternative infrastructures that marginalized users in high-stakes contexts use to bypass roadblocks they experience due to platform design and governance. This type of repair work is dynamic; the platforms change at the same time as their most vulnerable users, improving their design in tune with the needs of users. It will also enable users in high-stakes contexts to adopt alternative infrastructure designs, allowing interventions to scale to broader populations in moments of precarity, such as large-scale crisis events. This type of adoption of alternative infrastructure design could, however, be extractive, if unpaid and unacknowledged. More research is needed to understand how incentivization could compensate marginalized populations for their often invisible labor while also ensuring that bad-faith actors do

not take advantage of the same incentives. In addition, we advocate that collaborators and non-governmental organizations working at the intersection of social media and marginality partner with relevant stakeholders within communities experiencing precarity to fund and develop localized solutions that address the complex infrastructural needs and preserve their online presence.

While we advocate for platforms to collaborate with marginalized users to integrate alternative infrastructural features, our findings, lived experiences, substantial ongoing research, and the current platform governance climate suggest this outcome is fundamentally at odds with the platforms' prevailing logics of digital colonialism and efficiency. As Shahid and Vashistha [87] and Peterson-Salahuddin [75] argue, content moderation systems are designed to center Western norms and priorities, resulting in a coloniality of power that perpetuates historical injustices and systematically censors minoritized expressions. This governance model prioritizes punitive and efficient approaches over the safety and contextual needs of marginalized users [75], making platforms complicit in harm and structurally resistant to the deep infrastructural change required for true inclusivity. This tension presents an opportunity for future work that we hope that HCI researchers can engage with to understand how the structural and systemic challenges that hinder the implementation of platform design recommendations impact marginalized social media users experiencing precarity.

6 Conclusion

We presented findings from three case studies, diasporic groups in the Venezuelan humanitarian crisis, ethnic minorities in Nigeria facing violent attacks and human rights violations, and transgender people in the U.S. combating oppression, to demonstrate how the lived experiences of marginalized populations challenge the boundaries of existing social media platform design. Our research shows that these groups rely on social media as critical infrastructure to coordinate survival strategies, engage in activism, and amplify their voices. However, this essential use is constrained by rigid platform governance structures that fail to account for their urgent contextual needs.

Our core contributions show that: **(1)** For marginalized social media users, the breakdown of platform features is not a minor inconvenience. Still, it is symptomatic of systems that fail to perform in high-stakes contexts, rendering the complex needs of these users illegible to monolithic moderation infrastructures. **(2)** Platform failures force these users to engage in substantial, often invisible, sociotechnical repair work—including developing workarounds, self-censorship, and community governance, which ultimately reinforces their precarity. **(3)** The design gap in social media is rooted in decisions that perpetuate physical-world power asymmetries and digital colonialism, leaving marginalized populations as an afterthought.

To address the gap between what these marginalized populations need and what they experience, we recommend that platforms collaborate with them to integrate alternative infrastructural features, such as emergency response tools and exit mechanisms for wellbeing, which users are already devising. However, we also acknowledge the central paradox: that platforms, driven by prevailing logics of efficiency and digital colonialism, have little financial incentive to adopt these infrastructural changes that prioritize vulnerable users over punitive or uniform moderation. Future work must address these structural and systemic challenges to understand how the non-implementation of essential design recommendations continues to impact marginalized social media users experiencing precarity negatively.

Acknowledgments

We thank our study participants for sharing their valuable insights and experiences. Special thanks to our anonymous reviewers for their comments that improved this paper. We thank members of the Social Media Collective at Microsoft Research New England for their support of the data collection for Michaelanne's project (the first vignette). This research was supported by National Science Foundation grant 1942125 and National Science Foundation grant 2210841.

References

[1] Central Intelligence Agency. 2013. *The World Factbook 2012-13.* Central Intelligence Agency.
[2] Central Intelligence Agency. 2025. Nigeria - People and Society. https://www.cia.gov/the-world-factbook/countries/nigeria/#people-and-society Accessed: 2025-09-11.
[3] Kendra Albert and Oliver Haimson. 2022. Trans Content Moderation Scholars-Comments to Oversight Board. *Trans Content Moderation Scholars-Comments to Oversight Board (August 10, 2022)* (2022).
[4] Carolina Are. 2023. An autoethnography of automated powerlessness: lacking platform affordances in Instagram and TikTok account deletions. *Media, Culture & Society* 45, 4 (2023), 822–840.
[5] Carolina Are. 2024. 'Dysfunctional'appeals and failures of algorithmic justice in Instagram and TikTok content moderation. *Information, Communication & Society* (2024), 1–18.
[6] Anna Veronica Banchik. 2021. Disappearing acts: Content moderation and emergent practices to preserve at-risk human rights–related content. *New Media & Society* 23, 6 (2021), 1527–1544.
[7] Sarah Banet-Weiser. 2018. *Empowered: popular feminism and popular misogyny.* Duke University Press, Durham London.
[8] Reuben Binns, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. 2018. 'It's Reducing a Human Being to a Percentage' Perceptions of Justice in Algorithmic Decisions. In *Proceedings of the 2018 Chi conference on human factors in computing systems.* 1–14.
[9] Reuben Binns, Michael Veale, Max Van Kleek, and Nigel Shadbolt. 2017. Like trainer, like bot? Inheritance of bias in algorithmic content moderation. In *Social Informatics: 9th International Conference, SocInfo 2017, Oxford, UK, September 13-15, 2017, Proceedings, Part II 9.* Springer, 405–415.
[10] Sophie Bishop. 2019. Managing visibility on YouTube through algorithmic gossip. *New Media & Society* 21, 11-12 (Nov. 2019), 2589–2606. doi:10.1177/1461444819854731
[11] Tom Boellstorff. 2009. A typology of ethnographic scales for virtual worlds. In *Online worlds: Convergence of the real and the virtual.* Springer, 123–133.
[12] John Brewer. 2000. *Ethnography.* McGraw-Hill Education (UK).
[13] Catherine Buni and Soraya Chemaly. 2016. The secret rules of the internet. *The Verge* 13 (2016).
[14] Micha Cárdenas. 2022. *Poetic operations: Trans of color art in digital media.* Duke University Press.
[15] Juno Carmel. 2025. How Changes to U.S. Gender Policy Could Affect Your Passport. *The New York Times* (Feb. 2025). https://www.nytimes.com/2025/02/27/travel/trans-nonbinary-passport-changes-trump.html
[16] Antonio A. Casilli. 2017. Global Digital Culture| Digital Labor Studies Go Global: Toward a Digital Decolonial Turn. *International Journal of Communication* 11, 0 (Sept. 2017), 21. https://ijoc.org/index.php/ijoc/article/view/6349
[17] Marion G. Crain, Winifred Poster, and Miriam A. Cherry (Eds.). 2016. *Invisible labor: hidden work in the contemporary world.* University of California Press, Oakland, California.
[18] Arlene Kaplan Daniels. 1987. Invisible Work. *Social Problems* 34, 5 (Dec. 1987), 403–415. doi:10.2307/800538
[19] Dipto Das. 2023. Studying multi-dimensional marginalization of identity from decolonial and postcolonial perspectives. In *Companion Publication of the 2023*

Conference on Computer Supported Cooperative Work and Social Computing. 437–440.

[20] Dipto Das. 2024. *Identity Decolonization amid the Coloniality of Computing*. Ph. D. Dissertation. University of Colorado at Boulder.

[21] Dipto Das, Shion Guha, Jed R Brubaker, and Bryan Semaan. 2024. The"Colonial Impulse" of Natural Language Processing: An Audit of Bengali Sentiment Analysis Tools and Their Identity-based Biases. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–18.

[22] Giovanni De Gregorio and Nicole Stremlau. 2023. Inequalities and content moderation. *Global Policy* 14, 5 (2023), 870–879.

[23] Daniel Delmonaco, Samuel Mayworm, Hibby Thach, Josh Guberman, Aurelia Augusta, and Oliver L Haimson. 2024. " What are you doing, TikTok?": How Marginalized Social Media Users Perceive, Theorize, and" Prove" Shadowbanning. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–39.

[24] Thiago Dias Oliva. 2020. Content moderation technologies: Applying human rights standards to protect freedom of expression. *Human Rights Law Review* 20, 4 (2020), 607–640.

[25] Adio-Adet Tichafara Dinika. 2025. Navigating labour's labyrinth: Developing a typology of platform work in Sub-Saharan Africa. 2 (2025), 29768624251318084. doi:10.1177/29768624251318084 Publisher: SAGE Publications.

[26] Brooke Erin Duffy and Colten Meisner. 2023. Platform governance at the margins: Social media creators' experiences with algorithmic (in)visibility. *Media, Culture & Society* 45, 2 (March 2023), 285–304. doi:10.1177/01634437221111923 Publisher: SAGE Publications Ltd.

[27] Brooke Erin Duffy and Megan Sawey. 2022. In/Visibility in Social Media Work: The Hidden Labor Behind the Brands | Article | Media and Communication. (Jan. 2022). https://www.cogitatiopress.com/mediaandcommunication/article/view/4460

[28] Reshmi Dutt-Ballerstadt. 2020. In Our Own Words: Institutional Betrayals. (2020). https://www.insidehighered.com/advice/2020/03/06/underrepresented-faculty-members-share-real-reasons-they-have-left-various

[29] Michaelanne Dye. 2021. Un grano de arena: Infrastructural care, social media platforms, and the Venezuelan humanitarian crisis. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–28.

[30] Motahhare Eslami, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton, and Alex Kirlik. 2016. First I "like" it, then I hide it: Folk Theories of Social Feeds. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 2371–2382. doi:10.1145/2858036.2858494

[31] Megan French and Jeff Hancock. 2017. What's the Folk Theory? Reasoning About Cyber-Social Systems. *SSRN Electronic Journal* (2017). doi:10.2139/ssrn.2910571

[32] Florencia García-Rapp. 2017. Popularity markers on YouTube's attention economy: the case of Bubzbeauty. *Celebrity Studies* 8, 2 (April 2017), 228–245. doi:10.1080/19392397.2016.1242430 Publisher: Routledge _eprint: https://doi.org/10.1080/19392397.2016.1242430.

[33] Akriti Gaur. 2020. *Moderate Globally Impact Locally: Tackling Social Media's Hate Speech Problem in India*. https://law.yale.edu/moderate-globally-impact-locally-tackling-social-medias-hate-speech-problem-india

[34] Susan A. Gelman and Cristine H. Legare. 2011. Concepts and Folk Theories. *Annual Review of Anthropology* 40, 1 (Oct. 2011), 379–398. doi:10.1146/annurev-anthro-081309-145822

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

[36] Tarleton Gillespie. 2022. Do Not Recommend? Reduction as a Form of Content Moderation. *Social Media + Society* 8, 3 (July 2022), 20563051221117552. doi:10.1177/20563051221117552

[37] Zoe Glatt. 2022. We're all told not to put our eggs in one basket": uncertainty, precarity and cross-platform labor in the online video influencer industry. *International Journal of Communication* 16 (Jan. 2022), 3853 – 3871. https://ijoc.org/index.php/ijoc Num Pages: 19.

[38] Robert Gorwa, Reuben Binns, and Christian Katzenbach. 2020. Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society* 7, 1 (2020), 2053951719897945.

[39] Cheng Guo. 2020. Identity and User Behavior in Online Communities. In *Companion Proceedings of the 2020 ACM International Conference on Supporting Group Work (GROUP '20)*. Association for Computing Machinery, New York, NY, USA, 35–38. doi:10.1145/3323994.3371018

[40] Oliver L Haimson, Justin Buss, Zu Weinger, Denny L Starks, Dykee Gorrell, and Briar Sweetbriar Baron. 2020. Trans time: Safety, privacy, and content warnings on a transgender-specific social media site. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–27.

[41] Oliver L Haimson, Avery Dame-Griff, Elias Capello, and Zahari Richter. 2021. Tumblr was a trans technology: the meaning, importance, history, and future of trans technologies. *Feminist media studies* 21, 3 (2021), 345–361.

[42] Oliver L Haimson, Daniel Delmonaco, Peipei Nie, and Andrea Wegner. 2021. Disproportionate removals and differing content moderation experiences for conservative, transgender, and black social media users: Marginalization and moderation gray areas. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–35.

[43] Oliver L. Haimson and Anna Lauren Hoffmann. 2016. Constructing and enforcing "authentic" identity online: Facebook, real names, and non-normative identities. *First Monday* (June 2016). doi:10.5210/fm.v21i6.6791

[44] Oliver L Haimson, Kai Nham, Hibby Thach, and Aloe DeGuia. 2023. How transgender people and communities were involved in trans technology design processes. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.

[45] Camille Harris, Amber Gayle Johnson, Sadie Palmer, Diyi Yang, and Amy Bruckman. 2023. " Honestly, I Think TikTok has a Vendetta Against Black Creators": Understanding Black Content Creator Experiences on TikTok. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–31.

[46] Christine Hine. 2020. *Ethnography for the internet: Embedded, embodied and everyday*. Routledge.

[47] The White House. 2025. Defending Women From Gender Ideology Extremism And Restoring Biological Truth To The Federal Government. https://www.whitehouse.gov/presidential-actions/2025/01/defending-women-from-gender-ideology-extremism-and-restoring-biological-truth-to-the-federal-government/

[48] J Brian Houston, Joshua Hawthorne, Mildred F Perreault, Eun Hae Park, Marlo Goldstein Hode, Michael R Halliwell, Sarah E Turner McGowen, Rachel Davis, Shivani Vaid, Jonathan A McElderry, et al. 2015. Social media and disasters: a functional framework for social media use in disaster planning, response, and research. *Disasters* 39, 1 (2015), 1–22.

[49] Lara Houston, Steven J Jackson, Daniela K Rosner, Syed Ishtiaque Ahmed, Meg Young, and Laewoo Kang. 2016. Values in repair. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 1403–1414.

[50] Lilly C. Irani and M. Six Silberman. 2013. Turkopticon: interrupting worker invisibility in amazon mechanical turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. Association for Computing Machinery, New York, NY, USA, 611–620. doi:10.1145/2470654.2470742

[51] Steven J Jackson. 2014. 11 rethinking repair. *Media technologies: Essays on communication, materiality, and society* (2014), 221–39.

[52] Steven J Jackson, Alex Pompe, and Gabriel Krieshok. 2012. Repair worlds: maintenance, repair, and ICT for development in rural Namibia. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. 107–116.

[53] Sandy E. James, Jody L. Herman, Laura E. Durso, and Rodrigo Heng-Lehtinen. 2024. *Early Insights: A Report of the 2022 U.S. Transgender Survey*. Technical Report. National Center for Transgender Equality, Washington, DC.

[54] Kylie Jarrett. 2022. *Digital labor*. polity, Cambridge, UK Medford, MA.

[55] Sena A. Kojah, Kentaro Toyama, and Oliver L. Haimson. 2025. At Risk Marginalized Populations in High-Stakes Violence and Human Rights Contexts in Nigeria. *in press* (2025). doi:10.1145/3757504 School of Information, University of Michigan, USA.

[56] Sena A Kojah, Ben Zefeng Zhang, Carolina Are, Daniel Delmonaco, and Oliver L Haimson. 2025. " Dialing it Back:" Shadowbanning, Invisible Digital Labor, and how Marginalized Content Creators Attempt to Mitigate the Impacts of Opaque Platform Governance. *Proceedings of the ACM on Human-Computer Interaction* 9, 1 (2025), 1–22.

[57] Michael Kwet. 2019. Digital colonialism: US empire and the new imperialism in the Global South. *Race & class* 60, 4 (2019), 3–26.

[58] Kate Sangwon Lee. 2019. Explicit disaster response features in social media: safety check and community help usage on Facebook during Typhoon Mangkhut. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–12.

[59] Calvin A Liang, Sean A Munson, and Julie A Kientz. 2021. Embracing four tensions in human-computer interaction research with marginalized people. *ACM Transactions on Computer-Human Interaction (TOCHI)* 28, 2 (2021), 1–47.

[60] Renkai Ma and Yubo Kou. 2021. " How advertiser-friendly is my video?": YouTuber's Socioeconomic Interactions with Algorithmic Content Moderation. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–25.

[61] Renkai Ma, Yue You, Xinning Gui, and Yubo Kou. 2023. How Do Users Experience Moderation?: A Systematic Literature Review. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–30.

[62] Eliza Mackintosh. 2021. Facebook Knew It Was Being Used to Incite Violence in Ethiopia. It Did Little to Stop the Spread, Documents Show. *CNN* (25 Oct. 2021). https://www.cnn.com/2021/10/25/business/ethiopia-violence-facebook-papers-cmd-intl/index.html

[63] Gloria Mark and Bryan Semaan. 2009. Expanding a country's borders during war: the internet war diary. In *Proceedings of the 2009 international workshop on Intercultural collaboration*. 3–12.

[64] Adrienne Massanari. 2017. #Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. *New Media & Society* 19, 3 (March 2017), 329–346. doi:10.1177/1461444815608807 Publisher: SAGE Publications.

[65] Samuel Mayworm, Kendra Albert, and Oliver L Haimson. 2024. Misgendered During Moderation: How Transgender Bodies Make Visible Cisnormative Content Moderation Policies and Enforcement in a Meta Oversight Board Case. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. 301–312.

[66] Samuel Mayworm, Michael Ann DeVito, Daniel Delmonaco, Hibby Thach, and Oliver L Haimson. 2024. Content moderation folk theories and perceptions of platform spirit among marginalized social media users. *ACM Transactions on Social Computing* 7, 1-4 (2024), 1–27.

[67] Jaron Mink, Miranda Wei, Collins W Munyendo, Kurt Hugenberg, Tadayoshi Kohno, Elissa M Redmiles, and Gang Wang. 2024. It's Trying Too Hard To Look Real: Deepfake Moderation Mistakes and Identity-Based Bias. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–20.

[68] Paul Mozur. 2018. A Genocide Incited on Facebook, With Posts From Myanmar's Military. *The New York Times* (15 Oct. 2018). https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html

[69] Sarah Myers West. 2018. Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society* 20, 11 (2018), 4366–4383.

[70] Hellina Hailu Nigatu and Inioluwa Deborah Raji. 2024. "I Searched for a Religious Song in Amharic and Got Sexual Content Instead": Investigating Online Harm in Low-Resourced Languages on YouTube.. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. 141–160.

[71] Barbara Ortutay. 2025. Meta rolls back hate speech rules as Zuckerberg cites 'recent elections' as a catalyst. (2025). https://apnews.com/article/meta-facebook-hate-speech-trump-immigrant-transgender-41191638cd7c720b950c05f9395a2b49 Section: Business.

[72] Susanna Paasonen. 2021. *Dependent, distracted, bored: affective formations in networked media*. The MIT Press, Cambridge [Massachusetts].

[73] F Patel and L Hecht-Felella. 2021. Oversight Board's First Rulings Show Facebook's Rules Are a Mess. *Just Security* (2021).

[74] Billy Perrigo. 2019. Facebook Says It's Removing More Hate Speech Than Ever Before. But There's a Catch. *TIME* (26 Nov. 2019). https://time.com/5739688/facebook-hate-speech-languages/

[75] Chelsea Peterson-Salahuddin. 2024. Repairing the harm: Toward an algorithmic reparations approach to hate speech content moderation. *Big Data & Society* 11, 2 (2024), 20539517241245333.

[76] Ryan Plis and Evelyn Blackwood. 2012. Trans technologies and identities in the United States. In *Technologies of Sexuality, Identity and Sexual Health*. Routledge, 185–204.

[77] Jacob Poushter, Caldwell Bishop, and Hanyu Chwe. 2018. Social media use continues to rise in developing countries but plateaus across developed ones. *Pew research center* 22 (2018), 2–19.

[78] Yim Register, Izzi Grasso, Lauren N Weingarten, Lilith Fury, Constanza Eliana Chinea, Tuck J Malloy, and Emma S Spiro. 2024. Beyond Initial Removal: Lasting Impacts of Discriminatory Content Moderation to Marginalized Creators on Instagram. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–28.

[79] Yim Register, Lucy Qin, Amanda Baughan, and Emma S Spiro. 2023. Attached to "The algorithm": Making sense of algorithmic precarity on instagram. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.

[80] Christian Reuter, Stefan Stieglitz, and Muhammad Imran. 2020. Social media in conflicts and crises. *Behaviour & Information Technology* 39, 3 (2020), 241–251.

[81] Sarah T. Roberts. 2019. *Behind the Screen: Content Moderation in the Shadows of Social Media*. Yale University Press, New Haven, CT.

[82] Jeffrey Sablosky. 2021. Dangerous organizations: Facebook's content moderation decisions and ethnic visibility in Myanmar. *Media, culture & society* 43, 6 (2021), 1017–1042.

[83] Mohammed A Salih. 2024. Facebook's platform coloniality: At the nexus of political economy, nation-state's internal colonialism, and the political activism of the marginalized. *New media & society* 26, 10 (2024), 6141–6158.

[84] Sarita Schoenebeck, Oliver L Haimson, and Lisa Nakamura. 2021. Drawing from justice theories to support targets of online harassment. *new media & society* 23, 5 (2021), 1278–1300.

[85] Farhana Shahid. 2024. *Colonialism in Content Moderation Research: The Struggles of Scholars in the Majority World*. https://cdt.org/insights/colonialism-in-content-moderation-research-the-struggles-of-scholars-in-the-majority-world/

[86] Farhana Shahid, Mona Elswah, and Aditya Vashistha. 2025. Think outside the data: Colonial biases and systemic issues in automated moderation pipelines for low-resource languages. *arXiv preprint arXiv:2501.13836* (2025).

[87] Farhana Shahid and Aditya Vashistha. 2023. Decolonizing Content Moderation: Does Uniform Global Community Standard Resemble Utopian Equality or Western Power Hegemony?. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18.

[88] Ellen Simpson, Andrew Hamann, and Bryan Semaan. 2022. How to Tame "Your" Algorithm: LGBTQ+ Users' Domestication of TikTok. *Proc. ACM Hum.-Comput. Interact.* 6, GROUP (Jan. 2022), 22:1–22:27. doi:10.1145/3492841

[89] Cheryll Ruth R. Soriano, Earvin Charles Cabalquinto, and Joy Hannah Panaligan. 2021. Performing "digital labor bayanihan": strategies of influence and survival in the platform economy. 23, 57 (2021), 84–111. doi:10.1590/15174522-113027

[90] Dean Spade. 2015. *Normal life: Administrative violence, critical trans politics, and the limits of law*. Duke University Press.

[91] Ella Steen, Kathryn Yurechko, and Daniel Klug. 2023. You can (not) say what you want: Using algospeak to contest and evade algorithmic content moderation on TikTok. *Social Media+ Society* 9, 3 (2023), 20563051231194586.

[92] Nicole Stremlau. 2022. Online speech and offline violence: Reflections on the current violence in Ethiopia. *Global Media Journal-German Edition* 12, 2 (2022).

[93] Lucy Suchman. 2020. Agencies in technology design: Feminist reconfigurations. In *Machine ethics and robot ethics*. Routledge, 361–375.

[94] Pratyay Suvarnapathaki, Viral Shah, Saarthak Negi, and Nimmi Rangaswamy. 2025. 'The Boring and the Tedious': Invisible Labour in India's Gig-Economy. *arXiv preprint arXiv:2504.17697* (2025).

[95] Hibby Thach, Samuel Mayworm, Daniel Delmonaco, and Oliver Haimson. 2022. (In)visible moderation: A digital ethnography of marginalized users and content moderation on Twitch and Reddit. *New Media & Society* 26, 7 (2022), 4034–4055. doi:10.1177/14614448221109804 Publisher: SAGE Publications.

[96] Hibby Thach, Samuel Mayworm, Michaelanne Thomas, and Oliver L Haimson. 2024. Trans-centered moderation: Trans technology creators and centering transness in platform and community governance: Trans-centered moderation. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. 326–336.

[97] Kristen Vaccaro, Christian Sandvig, and Karrie Karahalios. 2020. " At the End of the Day Facebook Does What ItWants" How Users Experience Contesting Algorithmic Content Moderation. *Proceedings of the ACM on human-computer interaction* 4, CSCW2 (2020), 1–22.

[98] Jenny Waycott, Greg Wadley, Stefan Schutt, Arthur Stabolidis, and Reeva Lederman. 2015. The Challenge of Technology Research in Sensitive Settings: Case Studies in'ensitive HCI'. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*. 240–249.

[99] Anindya Dessi Wulansari, Arif Novianto, Yeremias Torontuan Keban, and Ari Hernawan. 2024. Hiding behind the platform: the myth of flexibility for gig workers in Indonesia. 32, 1 (2024), 22–40. doi:10.1080/0967828X.2023.2292101 Publisher: Routledge _eprint: https://doi.org/10.1080/0967828X.2023.2292101.

[100] James Yékú. 2022. Digital African Literatures and the Coloniality of Data. *Cambridge Journal of Postcolonial Literary Inquiry* 9, 3 (2022), 381–398.