

Decoding Viral Racism on TikTok: A Content Analysis and Interview Study to Inform Educational Interventions

This study used a social listening tool and qualitative methods to analyze audiovisual content and user responses related to anti-Black and anti-Asian racism on the TikTok social media platform. Findings revealed that approximately 12% of posts contained racist content, mainly in the form of racial microaggressions and microassaults. We also identified user strategies that promote the virality of racist posts and qualitatively different trends in racist content across anti-Black and anti-Asian posts. Interviews with TikTok users further confirmed these trends and highlighted the need for an educational intervention to address online racism. This work is currently informing an online tool to help users recognize, report, and reduce the dissemination of racist content on TikTok.

August 2023

Prepared by

Christina L. Scanlon, Jane Schuchert Walsh, Jeffrey Tienes, Karen Sowon, Alice Huguet, Daniel Klug, Tuangtip Klinbubpa-Neff, Brady Shore

The Collaboratory Against Hate: Research and Action Center at Carnegie Mellon University and the University of Pittsburgh aspires to develop and support innovative multidisciplinary, interdisciplinary, and cross-university research aimed at understanding how extremist hate is generated, how it circulates in online and real-life spaces, and how it polarizes society and provokes harmful and illegal acts, especially toward communities of color and other minoritized groups. We seek to develop effective interventions to inhibit every stage in the creation and growth of extremist hate groups and to minimize their destructive consequences.

www.collabagainsthate.org

Carnegie Mellon University



1. Objective

Technology is an omnipresent facet of youth’s formal and informal education. While adults shepherd youth through formal virtual learning environments, less is known about the types of direct and vicarious education youth are receiving as a byproduct of participation in informal online spaces (e.g., social media platforms, such as TikTok). For instance, 72% of youth aged 13-17 from marginalized racial groups reported experiencing hate-based harassment online in 2021.¹ Exposure to online racism matters not only for minoritized youth’s mental health,² but it also poses negative consequences for all youth because of the risk of influencing users’ implicit biases and social media engagement.^{3,4} Furthermore, younger social media users are largely unaware of their role in spreading racist sentiments on social media.^{5,6} In this study, we partnered with ViralMoment—a cutting-edge social listening tool—to answer the following questions:

1

In what forms do we see racist content on the TikTok platform?

2

Are there observable user patterns in racist content that support virality on the Tiktok platform? Do these patterns vary between anti-Black and anti-Asian content?

3

How do users interpret and explain racist content on the TikTok platform?

Findings are currently being used to inform an educational intervention designed to help youth identify and respond to online racism.

2. Theoretical Framework

Extremist and hateful messages propagate the TikTok social media platform.^{1,7} Extremists have found ways to spread hateful messages through attention-grabbing viral content (e.g., 'humorous' skits or commentary)⁸ or intentional manipulation of the TikTok algorithm (e.g., algospeak where letters are replaced with numbers, such as H8 = hate).⁹ TikTok is a particularly nefarious context for racism due to the way its algorithm generates suggested content based on a user's history of liking, sharing, creating, and commenting on posts. Concerningly, the TikTok algorithm is designed to steer users toward content based on past engagement; hence, even inadvertent or brief engagement with a covertly racist TikTok post has the potential to steer users toward additional racial content.¹⁰ By participating in TikTok, then, youth are inadvertently 'opting in' to an increased likelihood of exposure to racist content with the potential to (a) influence unconscious biases about marginalized and minoritized groups and (b) steer TikTok algorithms toward progressively more extreme content.

In their *Taxonomy of Online Racism*,¹¹ Tynes et al. describe online racism as a system of ideologies, stereotypes, and emotions that dehumanize People of Color in ways that give privileged power to White individuals and maintain racial hierarchies in digital spaces. This taxonomy delineates three categories of online racism: microaggressions, microassaults, and hate crimes. Online racial microaggressions involve subtle, often unintentional representations of racist ideas, while online racial microassaults refer to intentional verbal or visual assaults based on race. Online hate crimes encompass criminal activities that target individuals based on their race. This taxonomy has been used in literature to better understand the content youth are exposed to online and how it impacts child and adolescent development.¹² Despite the proliferation of racist content on the TikTok platform, little work has been done to understand how and why racist posts 'go viral' despite automated and user-generated content moderation.

3. Methods

Social listening is the process of monitoring social media to understand key themes, discussions, and patterns as they spread.¹³ The *ViralMoment* (VM) software works by pulling the most widely circulated TikTok posts using given hashtags and creating a searchable database—referred to as a dashboard—that includes user information, audio transcriptions, text on screen, and visual components (e.g., symbols, logos).

3.1 Data Sources

During spring 2020, there were noted rises in anti-Black and anti-Asian sentiments linked to the George Floyd murder and the COVID-19 pandemic, respectively. To investigate these trends, VM created two dashboards of TikTok posts likely to contain racist content using a list of hashtags surrounding anti-Black police brutality (i.e., #AllLivesMatter, #BlackLivesMatter, #BlueLivesMatter, #ColinKaepernick, #DerekChauvin, #GeorgeFloyd) and pandemic-related anti-Asian biases (i.e., #Asian #China, #COVID, #Oriental, #StopAsianHate, #Virus). Our deductive and inductive analysis of the anti-Black (n = 2,111 posts) and anti-Asian (n = 1,841 posts) dashboards then informed a semi-structured interview protocol. In Spring 2023, qualitative interviews were conducted with 35 active TikTok users. Questions asked participants about their understanding of potentially racist content encountered on the TikTok platform as well as their motivations for how they interact with such videos.

3.2 Analytic Plan

Content Analysis

Deductive Coding. In Summer 2022, VM generated two dashboards: one pertaining to anti-Black police brutality and one pertaining to anti-Asian rhetoric surrounding COVID-19. We then classified each of these posts based on the *Taxonomy of Online Racism* (Stewart et al., 2019) to better understand the types of racist content that are present in widely circulated posts on the TikTok platform.

Inductive Coding. In Fall 2022, we reviewed posts within the VM dashboards that were flagged as containing racism to identify TikTok user strategies that may have prevented the post from being removed by TikTok content moderation. We also examined the anti-Black and anti-Asian dashboards as to whether there were qualitatively different trends in content and/or strategies.

Interview Study

In Spring 2023, we conducted an interview study to examine TikTok users' (a) interpretations of online racism and (b) online behaviors related to creating, sharing, liking, or commenting on racist content. Interviews lasted between 20-55 minutes and were conducted via Zoom. After interviews were deidentified and transcribed, we used *Dedoose*, an online qualitative analysis program, to conduct open coding of the 35 interview responses.

4. Results

4.1 Content Analysis

Preliminary deductive analysis indicates that out of the 3,952 total dashboard videos posted between January 2020 and July 2022, approximately 12% contained audio, visual, and/or text content that fit under the definitions for microaggressions, microassaults, and hate crimes per Tynes et al.'s Taxonomy of Online Racism. Prominent forms of microaggressions included misinformation/disinformation, microinvalidations, and microinsults, while microassaults were observed in individual and vicarious attacks, reinforcement of stereotypes, and racial humor. Hate crimes were rarely observed, likely due to automated and user-generated content moderation; however, it is important to note that the same process was less effective at identifying and addressing posts containing racial microaggressions and microassaults.

Our inductive analysis took a deeper look at the 12% of posts containing microaggressions, microassaults, and hate crimes (as coded in the deductive analysis) to identify user strategies that promote virality or evade content moderation. Strategies included embedding racist text within benign images; proliferating racist content and misinformation across smaller, less prominent accounts; using algostrategies designed to evade text-based content moderation (i.e., the use of L337, algospeak, or intentional misspellings); and hijacking the virality of popular audiovisual content (e.g., sounds, symbols, challenges). At times, audio, visual, and text components of posts worked together in ways that any given element could not be classified as racist in isolation. Rather, the racist message emerged in the combination of audiovisual content, thus illustrating the challenges of content moderation on the TikTok platform.

Inductive coding found trends relating to humor, mis/disinformation, and incidences where people posted real-life encounters with racism. Although our preliminary analysis indicates a similar prevalence of racist posts across dashboards, anti-Black and anti-Asian racism emerged in qualitatively different ways. For instance, posts containing racial humor on the anti-Asian dashboard tended to be self-referential (e.g., an Asian content creator makes fun of an Asian stereotype, perhaps as a means of coping), while posts containing racial humor on the anti-Black dashboard were often from anonymous content creators (i.e., the users race could not be determined) and used racial slurs or derogatory comments as punchlines. Furthermore, mis/disinformation on the anti-Black dashboard tended to be more general in nature, with content largely related to whether systemic inequity in the U.S. (a) exists and (b) is a problem. Mis/disinformation on the anti-Asian dashboard was more specifically related to the COVID-19 pandemic. Lastly, posts containing cell phone videos of real-life interpersonal exchanges including racism were more prominent on the anti-Black dashboard.

4.2 Interview Analysis

Patterns identified in our content analysis were largely supported within participants' interview responses. Participants stated that racism on TikTok is mostly covert (vs. overt) or happening in comments (vs. within the post itself), and few participants described

encounters with overt racist content (i.e., hate crimes). When participants did see racist content on TikTok, they commented that it is typically posted from impersonal or meme accounts (i.e., accounts where the user does not post videos of themselves) and relies on the audiovisual interplay of visual, audio, and textual elements to deliver racial microaggressions or microassaults. Participants also commented on the role of humor in online racism, with many commenting on how stereotypes and prejudices often emerge within 'humor that harms' (e.g., jokes at the expense of a racial group) as opposed to 'humor that heals' (e.g., coping through self-referential humor). However, participants also saw many videos that were counteracting or pointing out other racist videos or were replies to racist videos, such as through the TikTok 'stitch' or 'duet' features.

Interviews also provided more information about actions TikTok users take in response to racist content. While some participants indicated more passive methods of dealing with racist content (i.e., scrolling past it, hitting 'not interested' to avoid similar content), others took a more active approach by alerting their TikTok community, blocking or reporting the account, or refuting racist content via comments and direct messages. Some participants felt that these actions helped reduce racist content on the TikTok platform; however, most felt that their engagement with racist content on TikTok would inadvertently add to the post's virality. For example, participants felt that the TikTok algorithm favors racist content because these posts often cleverly make use of viral trends and covert algorithm manipulation that make content moderation difficult. Lastly, participants who formally reported racist TikTok posts indicated that it was not clear how the post was reviewed or whether corrective action was taken. In sum, there was a general sentiment that not enough is being done to address racism on the TikTok platform or increase users' awareness of algorithmic manipulation as a means of supporting the virality of racist posts.

5. Significance and Conclusions

Although content moderation has become drastically more effective on text-based social media sites (i.e., Twitter, Reddit), the audiovisual nature of TikTok presents a challenge to traditional methods of social media listening. For instance, text embedded within an image or user-generated sounds typically will not trigger automated platform-initiated moderation. Since TikTok's unveiling in 2016, little empirical research has attended to these trends, perhaps due to the lack of tools to do so. In addition, we have found a shortage of literature addressing user interpretations of and engagement with racist posts on the TikTok platform. This study addressed both of these shortcomings by (a) using a social media listening program attuned to audiovisual elements to capture a snapshot of the virtual sociocultural context on the TikTok platform and (b) better understand user encounters and responses to racism within TikTok posts.

Our deductive content analysis found no shortage of racist content online, and our inductive analysis showed that TikTok users may be intentionally or unintentionally engaging in user practices that promote the virality and dissemination of racism. Our qualitative interviews largely confirmed findings within the content analysis and highlighted a distinct need for educational tools to help users identify and contend with online racism. Using this information, we are currently partnering with a technology company focused on mental health and education to design an online tool aimed to help users recognize, report, and reduce the proliferation of racist content on the TikTok platform. We will be pilot testing the intervention in Fall 2023, with preliminary results expected in Spring 2024.

References

1. Anti-Defamation League. *Online Hate and Harassment: The American Experience 2022.*; 2022. <https://www.adl.org/sites/default/files/pdfs/2022-09/Online-Hate-and-Harassment-Survey-2022.pdf>
2. Del Toro J, Wang MT. Online Racism and Mental Health Among Black American Adolescents in 2020. *J Am Acad Child Adolesc Psychiatry*. Published online 2022. doi:10.1016/j.jaac.2022.07.004
3. Serenko A, Turel O. A dual-attitude model of system use: The effect of explicit and implicit attitudes. *Information and Management*. 2019;56(5):657-668. doi:10.1016/j.im.2018.10.009
4. Turel O, Serenko A. Cognitive biases and excessive use of social media: The Facebook implicit associations test (FIAT). *Addictive Behaviors*. 2020;105. doi:10.1016/j.addbeh.2020.106328
5. Magsino I. Teens won't stop posting racist videos and challenges on TikTok; Experts explain why the problem continues. *Insider*. Published online May 11, 2020. <https://www.insider.com/tiktok-continues-to-have-problems-with-racist-videos-2020-5>
6. Holt B. Teens on TikTok have no clue they're perpetuating racist stereotypes. *Quartz*. Published online November 1, 2019. <https://qz.com/quartz/1738478/how-teens-on-tiktok-are-perpetuating-racist-stereotypes>
7. Weimann G, Masri N. Research Note: Spreading Hate on TikTok. *Studies in Conflict and Terrorism*. 2023;46(5):752-765. doi:10.1080/1057610X.2020.1780027
8. Matamoros-Fernández A, Rodríguez A, Wikström P. Humor That Harms? Examining Racist Audio-Visual Memetic Media on TikTok During Covid-19. *Media Commun*. 2022;10(2):180-191. doi:10.17645/mac.v10i2.5154
9. Klug D, Steen E, Yurechko K. How Algorithm Awareness Impacts Algospeak Use on TikTok. In: *ACM Web Conference 2023 - Companion of the World Wide Web Conference, WWW 2023*. Association for Computing Machinery, Inc; 2023:234-237. doi:10.1145/3543873.3587355
10. Noble SU. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press; 2018.
11. Tynes BM, Lozada FT, Smith NA, Stewart AM. From Racial Microaggressions to Hate Crimes: A Model of Online Racism Based on the Lived Experiences of Adolescents of Color. In: Torino GC, ed. *Microaggression Theory: Influence and Implications*. John Wiley & Sons, Inc.; 2019:194-211. <http://ebookcentral.proquest.com/lib/pitt-ebooks/detail.action?docID=5526226>
12. Stewart AM, Maxie-Moreman AD, Schuschke J, Lozada FT, Rodgers AJ. From chat rooms to classrooms and TikToks too: Dr. Brendesha Tynes' contributions to the study of race and development in digital contexts. *American Psychologist*. 2023;78(4):551-562. doi:10.1037/amp0001124
13. Stewart MC, Arnold CL. Defining Social Listening: Recognizing an Emerging Dimension of Listening. *International Journal of Listening*. 2018;32(2):85-100. doi:10.1080/10904018.2017.1330656