Disinformation and Hate Speech Harm BIPOC

The National Hispanic Media Coalition

Maria Mercado

I. Introduction

The National Hispanic Media Coalition (NHMC) is a 35-year old civil rights organization dedicated to increasing Latinx representation in media and entertainment, and protecting and expanding the digital rights of the Latinx community and other marginalized communities. The depiction of Latinx in the media affects the context in which we are viewed in society. It shapes our interactions and our empowerment to seek opportunities. Thus, curtailing the spread of disinformation and hate speech is critical to ensuring that Latinx are given the best possible chance to progress. It has become a safety issue for our Latinx community as we are under-resourced when it comes to digital and media literacy. NHMC is dedicated to eliminating hate online and holding social media companies accountable for their role in the rise of white supremacy.

II. Statement of Purpose

The purpose of this paper is to critically examine the correlation between rises in disinformation and hate speech and the increased suppression of people of color. In addition, this paper seeks to investigate whether governmental structures and platform community standards have become vehicles of inaction, adding to the adversity faced by Latinx, Black, Indigenous, and other people of color.

III. Background

To draw the nexus between how hate speech and disinformation impact the social, political, and economic framework of marginalized communities, this paper will track and comment on the changes that each area has experienced as a result of their rise in prevalence on social media platforms in recent years. Undoubtedly, disinformation and hate speech affect critical happenings in our society, which breeds a necessity to delve deeper into understanding of what is going on.

Social media is a relatively new concept. Having only been created a little more than 20 years ago, social media has given way to a new generation of individuals who are essentially born navigating and mastering how to connect on an unprecedented, global scale. With the rise of social media, more opportunities arise everyday to elevate and spread the many political and cultural narratives that coexist around the world and in the United States in particular. Because of the newness of social media platforms, their technology, and business models, much of what has taken place online has gone unrestricted and unregulated. The long term effects and real-life harms from the use of social media are still emerging. Social media content amplifies the existing cultural context of our real world, and even mirrors the hateful parts of our societies, giving voice to a plethora of perspectives–both common and extreme.

While we must strive for inclusivity in our society, we can not passively stand by and allow the dangerous and inflammatory nature of extremist and hate speech to threaten the safety of

marginalized communities online. In order to understand why we see an unprecedented level of hate online, we must first acknowledge our country's long history of hate. The United States has a hateful track record of dehumanizing communities of color, and allowing violence to persist against them, such as the excusal and erasure of mass lynchings, the displacement of indigenous communities, and the commercialization of Latinx culture.

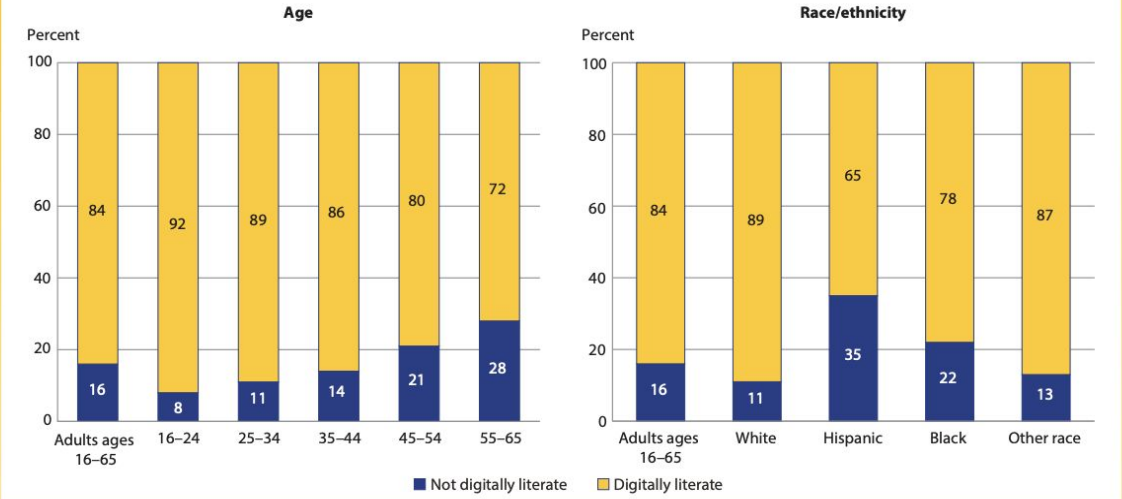
Today, social media has become a wary area, largely due to *viral culture*. Viral culture is the notion of when an event, post, or idea that quickly circulates on the internet, as well as the desire of social media users to achieve "vitality" through their content, often at questionable costs. Viral Culture makes it incredibly difficult for those who are looking to social media for reliable information and news. Platforms are saturated with questionable and false information, and consumers are having trouble deciding which sources to trust. This is compounded for Latinx, Black, Indigenous, and other people of color, who, historically, do not have access to technology, causing a gap in digital literacy skills.

Communities of color face a number of barriers in accessing information, making them more prone to believe posts on social media that are widely available for audiences to consume, despite their false or misleading nature. According to Figure 6, provided by the National Center for Education Statistics, Hispanics, followed by Black and "Other" people have the highest rate of digitally illiterate individuals (Pawlowski and Mamedova, 9). This, of course, stems from racial inequities, including the achievement gap, which refers to lower academic performance by students of color as a result of education inequities. BIPOC do not have the digital skills necessary to filter accurate or true knowledge and information from false, particularly when compared to white or more affluent counterparts.[1]

[1] Pawlowski, E., & Mamedova, S. (2018, May 29). A Description of U.S. Adults Who Are Not Digitally Literate. Retrieved August 20, 2020, from https://nces.ed.gov/pubs2018/2018161.pdf.

FIGURE 6.

DIGITAL LITERACY BY AGE AND RACE/ETHNICITY
Rate of digital literacy among U.S. adults ages 16–65, by age and by race/ethnicity: 2012



NOTE: *Other race* includes Asian, American Indian or Alaska Native, Hawaiian or other Pacific Islander, and persons of Two or more races. Race categories exclude persons of Hispanic ethnicity. Detail may not sum to 100 because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Organization for Economic Cooperation and Development (OECD), Program for the International Assessment of Adult Competencies (PIAAC), 2012.

Moreover, there are levels to online deceptive practices including misinformation, mal-information, and, the more prevalent, disinformation. These terms operate in a tiered hierarchy: misinformation is false information with no mal-intent of spread; mal-information is information that is based in truth spread with the intention of adversely affecting an individual or group; and disinformation is false information spread with the intention to deceive. These information practices can often involve real life issues, events which cause devastating effects on vulnerable populations if used to harm or with mal-intent.

Often on social media, where we see misinformation, disinformation, or mal-information, we also see hate speech. Hate speech, as a whole, is defined as “any form of expression through which speakers intend to vilify, humiliate, or incite hatred against a group or a class of persons on the basis of race, religion, skin color, sexual identity, gender identity, ethnicity, disability, or national origin” (Ward, 765).[2] Since 2016, and throughout Donald Trump’s Presidency, we have seen a frameshift change in the culture of America that has invited a sense of freedom into the minds of prejudiced Americans, allowing them to more freely extend hate to people without consequence. According to the FBI in 2018, hate crimes reached a 16-year high with the largest increase in crimes targeting the Latinx population (U. S. Federal Bureau of Investigation, 2018).[3] This

[2] Ward, K. D. (1998, April 1). Free Speech and the Development of Liberal Virtues: An Examination of the Controversies Involving Flag-Burning and Hate Speech. University of Miami School of Law Institutional Repository. https://repository.law.miami.edu/cgi/viewcontent.cgi?article=1693&context=umlr.
[3] U. S. Federal Bureau of Investigation Uniform Crime Reporting. (2018). Incidents, Offenses, Victims, and Known Offenders. https://ucr.fbi.gov/hate-crime/2018/tables/table-1.xls.

increase is the result of repeated assaults on the character of our community by Republican leadership and conservative public figures. For instance, the Latinx community is othered by popular referencing to Latinxs as “them” or “aliens”, Donald Trump’s infamous othering in 2016 where he grouped (Lee, 2015, para. 4)[4] vulnerable immigrants with “criminals, drug dealers, and rapists”,[5] and Republican Senators, like Tom Cotton, that make it a point to label Latinxs as national security threats in this country.[6]

Hate speech has transcended physicality into our digital lives, and has become an even more effective means of targeting and harassing individuals, often *en mass*. According to the Anti-Defamation League, 53 percent of Americans have experienced hate online, with 37 percent of those people experiencing what is identified as “severe hate” (AntiDefamation League, n.d., para. 16).[7] BIPOC communities face adversity both in person and online. Every week we see videos on popular platforms showing stories of Latinx street vendors being beaten, or of Latinx shoppers being harassed with calls to ‘go back where they came from.’ Social media has become another avenue for vulnerable communities to be targeted with extreme hate.

Further, it is NHMC’s belief that how individuals are portrayed on social media dictates popular opinion and how they are viewed in our real world. The way cyberhate affects the mental health and overall prosperity of individuals and communities that experience it is an area of civil rights policy that desperately needs more attention. To demonstrate the dire need for more protections against hate, a study was conducted by Pew Research Center found that, “Device[s] use[d] will lead to more social alienation, increased depression and less-fit people. Because it’s still relatively new, its dangers are not well understood yet.” (Anderson and Rainie, 2019).[8] Even as researchers develop an understanding and track evidence for how communities of online hate victims are harmed, there is still hesitation by platforms and governments to address those issues.

---

[4] Lee, M. (2015, July 08). Analysis | Donald Trump's false comments connecting Mexican immigrants and crime. Retrieved August 10, 2020, from https://www.washingtonpost.com/news/fact-checker/wp/2015/07/08/donald-trumps-false-comments-connecting-mexican-immigrants-and-crime/.

[5] Stop Illegal Immigration [Advertisement]. (2020, July 9). Retrieved 2020, from https://www.facebook.com/ads/library/?id=601993387118308.

[6] Tom Cotton for Senate [Advertisement]. (2020, July 24). Retrieved 2020, from https://www.facebook.com/ads/library/?id=2642522742676579.

[7] Anti-Defamation League (2019) Online Hate and Harassment: The American Experience. https://www.adl.org/onlineharassment.

[8]Anderson, J., & Rainie, L. (2019, December 31). Concerns about the future of people's well-being and digital life. Retrieved August 21, 2020, from https://www.pewresearch.org/internet/2018/04/17/concerns-about-the-future-of-peoples-well-being/.

IV. DURING THE COVID-19 PANDEMIC, DISINFORMATION RISKS THE SAFETY AND HEALTH OF COMMUNITIES OF COLOR.

Disinformation has significantly increased in volume on online platforms during the COVID-19 pandemic, and has led to negative health outcomes–particularly for people of color. Over the last several years, the world has become dependent on the internet for information and news. As we become more dependent on the internet for news, the opportunity for disinformation increases. According to the Oxford Internet Institute, disinformation campaigns in 2019 occurred in at least 70 countries where there were instances of disinformation in only 28 countries in 2017 (Bradshaw and Howard, 5).[9]

For communities that struggle with access to technology and digital skills training, disinformation is another barrier to civic engagement. Many disenfranchised communities lack the digital literacy and media literacy skills necessary to effectively sort through the abundant sources of information online. This is particularly true today, as we live amidst a global pandemic; filtering through information online could mean the difference between living or dying. With misinformation about the number and seriousness of COVID-19 cases, treatments for the virus, and the origins of the virus bombarding our online experiences, there is a natural mistrust between technology companies and consumers. Consumers do not feel protected on the online platforms they frequent, and bad actors are spreading propaganda and lies without facing any repercussions.

A. *The Trump Administration Is Making a Targeted Attempt to Spread Misinformation.*

Platforms, like Facebook, give politicians the ability to bypass community policies to post otherwise banned content and statements because they are considered newsworthy. COVID-19 makes this a more dangerous issue, as newsworthy policies and relaxed enforcement give politicians the opportunity to downplay the severity of the virus by lying about the numbers of coronavirus cases, encouraging anti-mask sentiment, spreading conspiracy theories about where the virus originated, and uplifting dangerous treatment with no basis in science. This continues to be Facebook's policy, despite it being widely unpopular. Sixty three percent of Americans believe that posts should be removed if they contain misinformation, and believe there needs to be a greater effort from social media platforms to curtail the spread of false information (Politico, 2).[10]

Only recently has the tech industry started to shift towards altruism in their content moderation practices. For example, after years of facing public scrutiny for allowing false information and

[9] Bradshaw, S., & Howard, P. (2020, August 04). Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation. Retrieved August 13, 2020, from http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/07/ct2018.pdf.
[10] Politico. (2020, May). Accountable Tech. https://www.politico.com/f/?id=00000172-5565-d57a-ad7b-5f6771540000.

statements made by President Trump to remain online, Twitter instituted a fact-checking feature earlier this year, where the platform verifies content as truth, marks it as questionable, or removes it if it is false. Twitter's move toward the truth has been met with harsh blowback from the Trump administration, leading many Americans to believe there is a clear intent by the Administration to deceive the public. He has demonstrated that even as an elected official, the moral obligation to be honest and truthful to constituents during times of emergency is lost.

Though Twitter has taken a stance against false information, other platforms like Facebook and Instagram, have not. This inaction is dangerous, as our healthcare and safety relies on the effective dissemination of truth. Forty-three percent of adults get news often from news websites or social media, according to Pew Research Center (Shearer, 2018, para. 3).[11] Our officials are exacerbating the impact and fatality rate of the COVID-19 pandemic, including its disproportionate effect on Black, Indigenous, and Latinx people, by using disinformation to deceive the public and gain political points.

To add, early panic during the pandemic brought on an onslaught of conspiracy theories that continue to be fabricated and shared by millions of people on the internet. For instance, Pew Research Center released a study that found that 29 percent of Americans believed that COVID-19 was developed intentionally in a lab (Jaiswal et. al, 1). [12] Information like this is dangerous for communities who already have difficulty trusting the government, like the Latinx community who mistrust government and authority, likely due to the amplification of anti-immigrant and anti-Latinx sentiment.

A perfect example of how disinformation campaigns can transcend into real life harms is in the Trump administration's spreading of false information about the origins of the coronavirus. By recklessly sharing that the novel coronavirus originated in China, President Trump and his administration ignited a further spike in xenophobia, in turn casting a severe lash against the Asian-American communities in the United States (Jaiswal et. al, 2).[13] This kind of false, dehumanizing, and hate-motivated commentary is viewed by millions on online platforms. Words, truth, and intent matter online, especially from an elected official. Online hatred and attacks have transitioned to physical assaults in public that then are exponentially shared. For instance, the case of an old man who was beaten and robbed by aggressors, one of which expressed that he "hates Asians," arose in March amidst the beginning of the pandemic, and

[11] Shearer, E. (2020, May 30). Social media outpaces print newspapers in the U.S. as a news source. Retrieved August 13, 2020, from https://www.pewresearch.org/fact-tank/2018/12/10/social-media-outpaces-print-newspapers-in-the-u-s-as-a-news-source/.

[12] Jaiswal, J., LoSchiavo, C., & Perlman, D. C. (2020, May 21). Disinformation, Misinformation and Inequality-Driven Mistrust in the Time of COVID-19: Lessons Unlearned from AIDS Denialism. AIDS and behavior.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7241063/pdf/10461_2020_Article_2925.pdf.

[13] Jaiswal, J., LoSchiavo, C., & Perlman, D. C. (2020, May 21). Disinformation, Misinformation and Inequality-Driven Mistrust in the Time of COVID-19: Lessons Unlearned from AIDS Denialism. AIDS and behavior. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7241063/pdf/10461_2020_Article_2925.pdf.

was shared and viewed by millions of people (Barmann, 2020, para. 2).[14] Another example, detailed in an article published by the New York Times, is of an encounter between a Chinese woman and a man who encouraged a bus to run her over and then proceeded to spit on her based on his belief that Asian Americans are to blame for the virus (Tavernise and Oppel, 2020).[15]

B. *Racist Histories Explain America's Reaction to COVID-19.*

BIPOC communities are often inclined to believe that the government is operating against them in some way or has plans to eradicate as many unassimilable people as possible. These aren't unfounded beliefs, as much of United States history is riddled with the erasure of culture and purification of society by imposing the elite's will on the inferior. This commentary serves as a link between the reality that the communities most vulnerable to disinformation harms are also the ones who share a troubled past with systems of power.

The idea that emergencies are the fault of a minority group is not new. During the Great Depression, many Americans channelled their insecurities into a group effort to deport Mexican-Americans in what is known as Mexican Repatriation. The rationale was that, "Mexican immigrants were supposedly using resources and working jobs that should go to White Americans affected by the Great Depression" (Little, 2019, para. 3).[16] Mexican-Americans, many of whom were legal residents, were ripped from their homes as sacrifices for something they were presumed to have caused. The recent rise in anti-Chinese sentiment is similar in the idea that many Americans are redirecting their fear into prejudice.

V. PRESIDENT TRUMP FUELS ANTI-IMMIGRANT SENTIMENT.

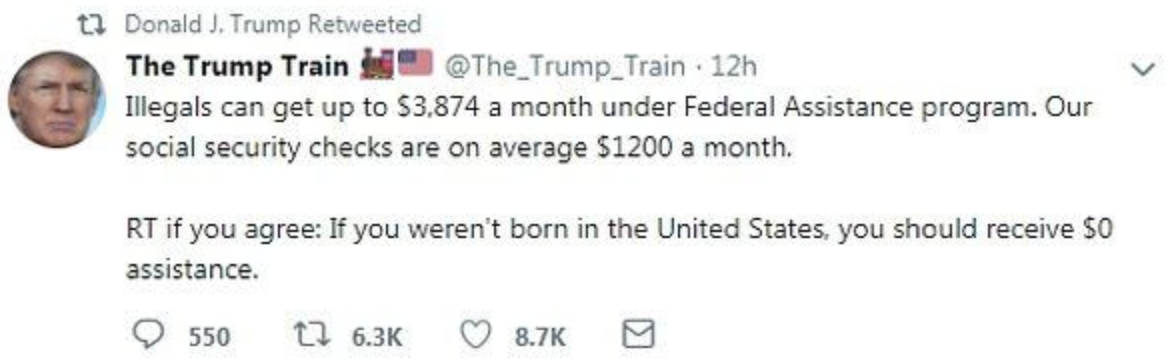
The past four years have been a never-ending period of disinformation, harassment, and hate for the Latinx community. President Trump and other conservative politicians and public figures have skillfully laid a path of stereotypes and xenophobia for their bases to inflate. For example, in an opinion article published by the New Yorker, Héctor Tobar responded to Trump's tweet below saying,

---

[14] Barmann, J. (2020, February 24). Disgusting Twitter Video Shows Apparent Bayview Robbery, Black-on-Asian Racism. Retrieved August 14, 2020, from https://sfist.com/2020/02/24/disgusting-twitter-video-shows-apparent-robbery-black-on-asian-racism/.

[15] Tavernise, S., & Oppel, R. A. (2020, March 23). Spit On, Yelled At, Attacked: Chinese-Americans Fear for Their Safety. The New York Times. https://www.nytimes.com/2020/03/23/us/chinese-coronavirus-racist-attacks.html.

[16] Little, B. (2019, July 12). The U.S. Deported a Million of Its Own Citizens to Mexico During the Great Depression. Retrieved August 18, 2020, from https://www.history.com/news/great-depression-repatriation-drives-mexico-deportation.



"There's a cottage industry now in the debunking of false facts circulating on the Internet, and very quickly the Washington Post, the Associated Press, and others revealed that the tweet was entirely incorrect" (Tobar, 2018, para. 4).[17] Nonetheless, Twitter allowed this post and others like it to persist on their platform. The amount of influence that President Trump holds is power that should be used thoughtfully and morally. Instead, it is being used to spread a false idea of what Latinx immigrants are–and social media platforms allow it. Latinxs endure targeted dehumanization as a result of the manipulation of information to convert us into "dangerous resource extractors."

The President has also used images to spread lies and hate about immigrants. In one viral instance, he used an image of people fighting to get across Morocco's border, falsely inferring that this is how Mexican immigrants are at the U.S. Southern border (Collins, 2016).[18] This image was far reaching, and largely has encouraged other politicians to follow suit to create a dehumanizing, homogenous picture of what the Mexican immigrant is. The immigrant and Latinx community suffer in opportunity, and face the burden of false information based on discriminatory stereotypes.

This kind of false information, rooted in hate, is not just about microaggressions and public sentiment; it's about the livelihood and survival of our people. The more that hate and disinformation grow, the more real-life manifestations occur. Disinformation is a weapon that is being yielded to stifle and silence the growing Latinx community, and it has had a long-lasting effect on how the community is viewed and valued.

A. *Disinformation Worsens Race Relations and Reinforces Trauma for the Black Community.*

[17] Tobar, H., & Gessen, M. (2018, December 12). Trump's Ongoing Disinformation Campaign Against Latino Immigrants. The New Yorker. https://www.newyorker.com/news/daily-comment/trumps-ongoing-disinformation-campaign-against-latino-immigrants.

[18] Collins, E. (2016, January 4). Trump ad uses footage from Morocco, not Mexican border. POLITICO. https://www.politico.com/story/2016/01/donald-trump-ad-footage-border-morocco-217332.

Disinformation about identity-based events has worsened race relations in the United States, and has manifested into real-world conflict. Social media has become less about connecting with loved ones, and instead, more about fulfilling agendas to target certain groups. Disinformation serves as a vehicle for harassment and defamation when in the hands of people online. Those actions have tangible effects on the lives of real individuals, and can cause irreparable harm.

Disinformation is also incredibly harmful to racial justice movements, including the rise in the movement for Black Lives in 2020, following the murder of George Floyd. Conspiracies continue to run rampant, suggesting that George Floyd is not dead, or that billionaires supplied bricks to protestors (Alba, 2020).[19] These conspiracy theories threaten to discredit or demoralize the racial justice movement, and are only some of the troubling fake stories being shared online for many to see (Guynn, 2020).[20]

Targeted, manipulative disinformation like this hurt people of color on the front lines, who are already suffering and exhausted with injustice. The real trauma from witnessing the murder of George Floyd, and so many others, at the hands of the police, is belittled by those who find comedy or power in false information, and reinforces centuries of affliction for the Black community. Race relations in this country suffer because of false narratives that are intended to further the divide between us.

B. Bad Actors Make a Targeted Effort to Influence Political Outcomes.

If effectively used, social media is a center for politicians and constituents to extend their political opinions and garnish support. One on hand, it gives elected officials the opportunity to reach out to constituencies that they would otherwise have trouble communicating with. However, consequently, some politicians have learned how to invent a narrative and disseminate that narrative online with the intent to manipulate public opinion.

In one particular instance, a video of former Vice President and 2020 presidential candidate Joe Biden was misleadingly edited to make him sound as if he was making racist remarks. The video showed Biden speaking about the culture of the United States, and how the culture needs to be changed. This video was spread across social media by right-wing media sources (Glueck, 2020),[21] and became more accessible than the original speech, which led to the

[19] Alba, D. (2020, June 01). Misinformation About George Floyd Protests Surges on Social Media. Retrieved August 14, 2020, from https://www.nytimes.com/2020/06/01/technology/george-floyd-misinformation-online.html.
[20] Guynn, J. (2020, June 2). George Floyd protests: How to avoid disinformation and misinformation on Facebook and Twitter. USA Today. https://www.usatoday.com/story/tech/2020/06/01/george-floyd-protests-disinformation-misinformation-surging-online/5313920002/.
[21] Glueck, K. (2020, January 3). Biden Warns About Disinformation After Misleading Video. The New York Times. https://www.nytimes.com/2020/01/02/us/politics/joe-biden-culture-video.html.

suppression of the truth. Users with low media and digital literacy, who are often people of color, likely would have a hard time determining whether this kind of false information is trustworthy.

Another example of disinformation used for political influence and manipulation is the Trump administration's ongoing anti-vote-by-mail campaign. In a news conference, he addressed the topic by saying, "universal mail-in voting is going to be catastrophic, it's going to make our country the laughing stock of the world" (BBC, 2020, para. 12).[22] According to ABC News, "49% of Americans are convinced mail-in voting is susceptible to significant fraud" (Karson and Cunningham, 2020, para. 1),[23] despite there being virtually zero data to support this notion (Kamarck and Stenglein, 2020).[24] An overwhelming push by conservatives on this topic clouds the integrity of our vote-by-mail system on the premise that if voting becomes more accessible for the average individual then power will be taken out of their hands. In essence, disinformation about voting by mail is playing a role in the disenfranchisement of BIPOC communities, and could lead to a suppression of their votes in November 2020.

C. Latinx Groups are the Most Vulnerable to Deception.

Digital literacy is among the one of the biggest factors of what makes the Latinx community, in particular, vulnerable to deception and manipulation by way of disinformation campaigns. According to the National Center for Education Statistics (NCES), Latinxs have the lowest rate of digital literacy relative to other races: 65 percent of Latinx adults were digitally literate while an average of 84 percent of all adults were digitally literate (Pawlowski and Mamedova, 8).[25] A study conducted by Manjul Shrestha of West Virginia University analyzed how the Latinx community has been targeted by misinformation through the media, finding that the majority of information pushed to the Latinx community consists of "one hit wonders", meaning content pushed out by users who are primarily celebrities, have massive followings, and are pro-Trump. One hit wonders receive the highest levels of engagement relative to how infrequently they post. Given that there are few media outlets that specifically target the Latinx community, the data suggests "that political trolls overall appear to have much more sophisticated techniques for targeting Latinxs and creating engaging content" (Shrestha, 39).[26] These users feed propaganda to the Latinx community on social media and, as we have already identified, Latinx

---

[22] BBC. Trump says universal mail-in voting would be 'catastrophic'. (2020, August 16). Retrieved August 18, 2020, from https://www.bbc.com/news/world-us-canada-53795876.

[23] Karson, K., & Cunningham, M. (2020, July 21). 'I don't trust it': Is Trump's false rhetoric on vote-by-mail resonating? Retrieved August 14, 2020, from https://abcnews.go.com/Politics/trust-trumps-false-rhetoric-vote-mail-resonating/story?id=71887848.

[24] Kamarck, E., & Stenglein, C. (2020, June 11). Low rates of fraud in vote-by-mail states show the benefits outweigh the risks. Retrieved August 19, 2020, from https://www.brookings.edu/blog/fixgov/2020/06/02/low-rates-of-fraud-in-vote-by-mail-states-show-the-benefits-outweigh-the-risks/

[25] Pawlowski, E., & Mamedova, S. (2018, May 29). A Description of U.S. Adults Who Are Not Digitally Literate. Retrieved August 20, 2020, from https://nces.ed.gov/pubs2018/2018161.pdf.

[26] Shrestha, M. (2018). Midterm 2018 and targeting Latino community through misinformation and disinformation online. The Research Repository @ WVU. https://researchrepository.wvu.edu/cgi/viewcontent.cgi?article=4746&context=etd.

people are often the least equipped to combat these tactics.

VI. CONTENT MODERATION IS BIASED AGAINST BIPOC.

Hate Speech online is targeting people of color in a disproportionate manner. Researchers found that leading AI models for processing hate speech were one-and-a-half times more likely to flag tweets as offensive or hateful when they were written by African Americans, and 2.2 times more likely to flag tweets (Ghaffray, 2019, para. 2)[27] written in African American English, which describes the "English that is primarily, but not exclusively, associated with the speech of African Americans" (Zienkiewicz, 2008, para. 1).[28] The creators of the software responsible for moderating content for each platform have put their inherent biases into the AI algorithms causing the programs to operate using racist protocols (Ghaffray, 2019, para.).[29] Therefore, people of color–who are largely underrepresented in engineering roles responsible for this kind of AI–are being enforced against as the main sources of hate speech, simply because of racial biases input into AI systems by programmers. Without adequate human quality and bias control, people of color are more likely to be kicked off these platforms or have their posts taken down based solely on their perceived race.

While social media companies claim that they have implemented an effective means of censorship that ensures hateful content will be taken down (Vega, 2020),[30] they are unequally doing so. ProPublica explains that, "[t]heir work amounts to what may well be the most far-reaching global censorship operation in history. It is also the least accountable: Facebook does not publish the rules it uses to determine what content to allow and what to delete" (Angwin, 2017, para. 18).[31] Facebook may be doing some of the necessary work to monitor hate speech, however, their operating systems are disproportionately and selectively enforcing against users of color. Essentially, Facebook's enforcement of its hate speech policies avoid serving communities in most need of having their voice heard.

---

[27] Ghaffary, S. (2019, August 15). The algorithms that detect hate speech online are biased against black people. Retrieved August 14, 2020, from https://www.vox.com/recode/2019/8/15/20806384/social-media-hate-speech-bias-black-african-american-facebook-twitter.

[28] Zienkiewicz, S. (2008). African American Vernacular English (AAVE). Retrieved August 14, 2020, from https://www.pdx.edu/multicultural-topics-communication-sciences-disorders/african-american-vernacular-english-aave.

[29] Ghaffary, S. (2019, August 15). The algorithms that detect hate speech online are biased against black people. Vox. https://www.vox.com/recode/2019/8/15/20806384/social-media-hate-speech-bias-black-african-american-facebook-twitter.

[30] Vega, N. (2020, June 26). Zuckerberg says Facebook will crack down on hate speech as ad boycott widens. Retrieved August 14, 2020, from https://nypost.com/2020/06/26/facebook-twitter-stocks-plummet-after-unilever-joins-ad-boycott/.

[31] Angwin,J., P. P. (2017, June 28). Facebook's Secret Censorship Rules Protect White Men From Hate Speech But Not Black Children. ProPublica. https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms.

A. Big Tech Protects Discrimination.

Action and inaction by platforms, like Facebook, concentrate resources to protect and reinforce discrimination instead of ensuring the safety of users most affected by hate and harassment. Bad actors are able to bypass the content moderation system with a number of nuanced tricks and loopholes designed with complacency by the platforms themselves. For example, a user can get by remarking that "...migrants can be referred to as 'filthy' but not called 'filth,' [so] they cannot be likened to filth or disease when the comparison is in the noun form (Angwin, 2017, para. 49)."[32] The nuances in the platforms' content moderation policies that serve to protect bad actors are reflective of the way that communities of color are often treated in the real world.

Through inaction, social media platforms are engaging in a concerted effort to silence people of color on their platforms by over censoring and safeguarding violent white-supremacist organizations. Jessica Guynn, a writer at USA Today, references a study by the Anti-Defamation League surrounding hate speech to clarify that, "[t]he reverberations can linger long after online attacks. Thirty-eight percent of the individuals surveyed who experienced online hate or harassment said they curtailed or changed their online habits. While 18 percent tried to contact the social media platform, 15 percent took steps to protect themselves and 6 percent contacted the police to ask for help or report the harassment" (Guynn, 2019, para 10).[33] In a time where more that half of Americans have been harassed online,[34] the onus is on social media companies or government officials to act or regulate in a way that works to protect and advance the civil rights of marginalized communities.

The prevalence of hate speech and the emotions that it elicits are unfair to the communities it targets, who already experience an undue burden. This is particularly true for women and women of color. For instance, a men's right-wing group can alter a picture of a woman's face to be bloody and bruised, and get away with posting it, because it does not violate Facebook's community standards (Buni and Chemaly, 2014).[35] Even when directly confronted, Facebook proceeds to make excuses about why clearly violent and dehumanizing content targeting women and women of color is not taken down.

---

[32] Angwin, J., P. P. (2017, June 28). Facebook's Secret Censorship Rules Protect White Men From Hate Speech But Not Black Children. ProPublica. https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms.

[33] Guynn, J. (2019, February 13). If you've been harassed online, you're not alone. More than half of Americans say they've experienced hate. USA Today. https://www.usatoday.com/story/news/2019/02/13/study-most-americans-have-been-targeted-hateful-speech-online/2846987002/.

[34] Guynn, J. (2019, February 13). If you've been harassed online, you're not alone. More than half of Americans say they've experienced hate. USA Today. https://www.usatoday.com/story/news/2019/02/13/study-most-americans-have-been-targeted-hateful-speech-online/2846987002/.

[35] Buni, C., & Chemaly, S. (2014, October 9). The Unsafety Net: How Social Media Turned Against Women. The Atlantic. https://www.theatlantic.com/technology/archive/2014/10/the-unsafety-net-how-social-media-turned-against-women/381261/.

Facebook has become a hub where extremists and violent individuals can speak freely, and be free from repercussions. It is beyond troubling that men can use the platform and project misogynist comments like, “women are like grass, they need to be beaten/cut regularly” or “you just need to be raped” without fearing that Facebook will identify their posts, and suspend their accounts. Women who are enduring violence that the World Health Organization deemed “a global health problem of epidemic proportions” are undoubtedly affected by their characterization on pages like these and by Facebook’s unwillingness to protect their safety.

VII. Conclusion

Disinformation and hate speech have disparate impacts on racial, ethnic, gender, and other marginalized populations, jeopardize the livelihood of those groups. This culture of hate and manipulation on social media has a chilling effect on marginalized communities driven by fear. Disproportionate enforcement of safety and outright inaction by platforms give no solace to people of color, as their online rights are not taken seriously. White supremacists are taking note of this, and using it to slander and abuse individuals in a covert fashion. Moving forward, more legislation, regulation, and corporate responsibility are needed to force social media companies to become more accountable for the content that is posted on their platforms. In addition, social media companies need to continue to feel the pressure of the public and of civil rights organizations to ensure that they fight to protect vulnerable communities.

Acknowledgements

I would like to express deep gratitude to the staff at the National Hispanic Media Coalition who gave me the opportunity to research this topic. They enabled me with the skillset to comprehensively look into the topic and experience the advocacy regarding it firsthand. More specifically, I am grateful to Daiquiri Ryan who provided me with invaluable guidance about the researching process and worked with me in the refining of this paper to ensure that this work could be something that I am proud of. She offered me unlimited availability to discuss and work through all of my findings. Additionally, I would like to thank Univision Communications, Inc. who made possible, as a first-year undergraduate, to work in a position that I would otherwise be overlooked and gave me the autonomy to do system disrupting work. Concludingly, I am humbled to work under the leadership of Brenda Castillo who fearlessly directs our organization and values the experience and knowledge that younger generations bring to the table.

References

Alba, D. (2020, June 01). Misinformation About George Floyd Protests Surges on Social Media. Retrieved August 14, 2020, from https://www.nytimes.com/2020/06/01/technology/george-floyd-misinformation-online.html .

Anderson, J., & Rainie, L. (2019, December 31). Concerns about the future of people's well-being and digital life. Retrieved August 21, 2020, from https://www.pewresearch.org/internet/2018/04/17/concerns-about-the-future-of-peoples-well-being/.

Angwin,J., P. P. (2017, June 28). Facebook's Secret Censorship Rules Protect White Men From Hate Speech But Not Black Children. ProPublica. https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms.

Anti-Defamation League.(2019). *Online Hate and Harassment: The American Experience*. https://www.adl.org/onlineharassment.

Barmann, J. (2020, February 24). Disgusting Twitter Video Shows Apparent Bayview Robbery, Black-on-Asian Racism. Retrieved August 14, 2020, from https://sfist.com/2020/02/24/disgusting-twitter-video-shows-apparent-robbery-black-on-asian-racism/.

BBC. Trump says universal mail-in voting would be 'catastrophic'. (2020, August 16). Retrieved August 18, 2020, from https://www.bbc.com/news/world-us-canada-53795876.

Bradshaw, S., & Howard, P. (2020, August 04). Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation. Retrieved August 13, 2020, from http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/07/ct2018.pdf.

Buni, C., & Chemaly, S. (2014, October 9). The Unsafety Net: How Social Media Turned Against Women. The Atlantic. https://www.theatlantic.com/technology/archive/2014/10/the-unsafety-net-how-social-media-turned-against-women/381261/.

Collins, E. (2016, January 4). Trump ad uses footage from Morocco, not Mexican border. POLITICO. https://www.politico.com/story/2016/01/donald-trump-ad-footage-border-morocco-217332.

Ghaffary, S. (2019, August 15). The algorithms that detect hate speech online are biased against black people. Vox.

https://www.vox.com/recode/2019/8/15/20806384/social-media-hate-speech-bias-black-african-american-facebook-twitter.

Glueck, K. (2020, January 3). Biden Warns About Disinformation After Misleading Video. The New York Times. https://www.nytimes.com/2020/01/02/us/politics/joe-biden-culture-video.html.Guynn, J. (2019, February 14). *If you've been harassed online, you're not alone. More than half of Americans say they've experienced hate*. USA Today. https://www.usatoday.com/story/news/2019/02/13/study-most-americans-have-been-targeted-hateful-speech-online/2846987002/.

Guynn, J. (2019, February 13). If you've been harassed online, you're not alone. More than half of Americans say they've experienced hate. USA Today. https://www.usatoday.com/story/news/2019/02/13/study-most-americans-have-been-targeted-hateful-speech-online/2846987002/.

Jaiswal, J., LoSchiavo, C., & Perlman, D. C. (2020, May 21). Disinformation, Misinformation and Inequality-Driven Mistrust in the Time of COVID-19: Lessons Unlearned from AIDS Denialism. AIDS and behavior.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7241063/pdf/10461_2020_Article_2925.pdf.

Kamarck, E., & Stenglein, C. (2020, June 11). Low rates of fraud in vote-by-mail states show the benefits outweigh the risks. Retrieved August 19, 2020, from https://www.brookings.edu/blog/fixgov/2020/06/02/low-rates-of-fraud-in-vote-by-mail-states-show-the-benefits-outweigh-the-risks/.

Karson, K., & Cunningham, M. (2020, July 21). 'I don't trust it': Is Trump's false rhetoric on vote-by-mail resonating? Retrieved August 14, 2020, from https://abcnews.go.com/Politics/trust-trumps-false-rhetoric-vote-mail-resonating/story?id=71887848.

Lee, M. (2015, July 08). Analysis | Donald Trump's false comments connecting Mexican immigrants and crime. Retrieved August 10, 2020, from https://www.washingtonpost.com/news/fact-checker/wp/2015/07/08/donald-trumps-false-comments-connecting-mexican-immigrants-and-crime/.

Little, B. (2019, July 12). The U.S. Deported a Million of Its Own Citizens to Mexico During the Great Depression. Retrieved August 18, 2020, from https://www.history.com/news/great-depression-repatriation-drives-mexico-deportation.

Pawlowski, E., & Mamedova, S. (2018, May 29). A Description of U.S. Adults Who Are Not Digitally Literate. Retrieved August 20, 2020, from https://nces.ed.gov/pubs2018/2018161.pdf.

Politico. (2020, May). Accountable Tech. https://www.politico.com/f/?id=00000172-5565-d57a-ad7b-5f6771540000.

Shearer, E. (2020, May 30). Social media outpaces print newspapers in the U.S. as a news source. Retrieved August 13, 2020, from https://www.pewresearch.org/fact-tank/2018/12/10/social-media-outpaces-print-newspapers-in-the-u-s-as-a-news-source/

Shrestha, M. (2018). Midterm 2018 and targeting Latino community through misinformation and disinformation online. The Research Repository @ WVU. https://researchrepository.wvu.edu/cgi/viewcontent.cgi?article=4746&context=etd.

Stop Illegal Immigration [Advertisement]. (2020, July 9). Retrieved 2020, from https://www.facebook.com/ads/library/?id=601993387118308.

Tavernise, S., & Oppel, R. A. (2020, March 23). Spit On, Yelled At, Attacked: Chinese-Americans Fear for Their Safety. The New York Times.Tobar, H., & Gessen, M. (2018, December 12). *Trump's Ongoing Disinformation Campaign Against Latino Immigrants*. The New Yorker. https://www.newyorker.com/news/daily-comment/trumps-ongoing-disinformation-campaign-against-latino-immigrants.

Tom Cotton for Senate [Advertisement]. (2020, July 24). Retrieved 2020, from https://www.facebook.com/ads/library/?id=2642522742676579.

U. S. Federal Bureau of Investigation Uniform Crime Reporting. (2018). Incidents, Offenses, Victims, and Known Offenders.https://ucr.fbi.gov/hate-crime/2018/tables/table-1.xls.

Vega, N. (2020, June 26). Zuckerberg says Facebook will crack down on hate speech as ad boycott widens. Retrieved August 14, 2020, from https://nypost.com/2020/06/26/facebook-twitter-stocks-plummet-after-unilever-joins-ad-boycott/.

Ward, K. D. (1998, April 1). Free Speech and the Development of Liberal Virtues: An Examination of the Controversies Involving Flag-Burning and Hate Speech. University of Miami School of Law Institutional Repository. https://repository.law.miami.edu/cgi/viewcontent.cgi?article=1693&context=umlr.

Zienkiewicz, S. (2008). African American Vernacular English (AAVE). Retrieved August 14, 2020, from

https://www.pdx.edu/multicultural-topics-communication-sciences-disorders/african-american-vernacular-english-aave.