WHEN ALGORITHMS SEE US: AN ANALYSIS OF BIASED CORPORATE SOCIAL MEDIA ALGORITHM PROGRAMMING AND THE ADVERSE EFFECTS THESE SOCIAL MEDIA ALGORITHMS CREATE WHEN THEY RECOMMEND HARMFUL CONTENT TO UNWITTING USERS

Sikudhani Foster- McCray*

CONTENTS

*Sikudhani Foster- McCray is a JD candidate at Emory University School of Law. Foster- McCray completed a B.A. in Business Management and Analytics in 2021 at Loyola University of New Orleans. Foster- McCray acknowledges Dr. Safiya Umoja Noble for her trailblazing efforts in the study of the intersectionality between artificial intelligence and race-based discrimination. Foster-McCray further extends her thanks to Dr. Ifetayo Ojelade from A Healing Paradigm Psychology Office for her interview contributions to this Paper.

WHEN ALGORITHMS SEE US: AN ANALYSIS OF BIASED CORPORATE SOCIAL MEDIA ALGORITHM PROGRAMMING AND THE ADVERSE EFFECTS THESE SOCIAL MEDIA ALGORITHMS CREATE WHEN THEY RECOMMEND HARMFUL CONTENT TO UNWITTING USERS

Sikudhani Foster- McCray

This Paper will analyze the injurious effects of racially biased algorithmic recommendation patterns on black communities. The analysis will build on the foundations of oppressive algorithm scrutiny, as presented in the Algorithms of Oppression text by Dr. Safiya Umoja Noble. Further, this Paper will indicate the current detrimental recommendation practices carried out by social media algorithms on platforms such as Facebook, Instagram, and Twitter. Upon this premise, this Paper will assess the legal implications of the biased algorithmic recommendations. Finally, this Paper will explore possible future legal claims available to harmed groups, against social media corporations for the unchecked actions of their algorithms.

INTRODUCTION

Algorithms are built into many programs, including those that transmit infinite databases of information and files for human consumption.[1] Some prominent platforms that run on the millions of rules comprising algorithms, include Google's search engine, Instagram's "Reels"[2] function, and Twitter's "Suggested tweets."[3] These platforms all work from rules programmed into billions of lines of rapidly operating computer code.[4] Because the code and algorithmic systems are incapable of self-critique for racial bias, an issue arises when the algorithms make unchecked choices, based on bias deeply embedded into their code.

Often, the humans who write the rules and code that comprise the algorithms are also often unaware and ill-equipped to assess, diagnose, and rework their own racial biases to stem the harmful effects of their own conscious and subconscious choices.[5]

According to internet, African- American, and gender studies professor Dr. Safiya Noble, "Human beings are developing the digital platforms we use, and as I present evidence of the recklessness and lack of regard that is often shown to

---

[1] Adam Mosseri, *Shedding More Light on How Instagram Works*, ANNOUNCEMENTS (June 8, 2021), https://about.instagram.com/blog/announcements/shedding-more-light-on-how-instagram-works.
[2] *Id.*
[3] *Our Approach To Recommendations*, HELP CENTER, https://help.twitter.com/en/rules-and-policies/recommendations.
[4] Mosseri, *supra* note 1.
[5] SAFIYA UMOJA NOBLE, ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM 2 (2018).

women and people of color in some of the output of these systems, it will become increasingly difficult for technology companies to separate their systematic and inequitable employment practices, and the far-right ideological bents of some of their employees, from the products they make for the public."[6] Noble wrote of this "recklessness and lack of regard"[7] from corporate employees, based on her research spanning from over a decade ago. Noble described both early-stage discriminatory practices, including hiring practices and algorithm programming by racially homogenous employees, and final stage practices, including corporate refusals to fix inaccurate algorithms and to remove racist, harmful content from search platforms.[8] Since Noble released her text, that focused on Google Search algorithms, other information sharing platforms have risen to prominence; including Facebook, Instagram, and Twitter.[9]

This Paper will expand on Noble's inquiry; pivoting from basic Google Search algorithmic discrimination, to incorporate an examination of dominant social media algorithmic discrimination. These newer algorithms not only produce a result when prompted by a user, but also independently suggest content to users unprompted.[10] An analysis of this novel feature is especially important because it will indicate how algorithms are being programmed to make connections and possibly racist assumptions about content, user desires, and promotion prioritizations. This Paper will critically analyze Noble's arguments in *Algorithms of Oppression*[11], propose a contemporary thesis concerning Google's social media counterparts, assess the current legal issues these social media platform algorithms have created, and finally, suggest future possible legal consequences social media corporations should face.

I. PREVIOUS WORK ON RACIALLY BIASED ALGORITHMS: *ALGORITHMS OF OPPRESSION*, BY DR. SAFIYA UMOJA NOBLE

Previous research and texts exist on cutting edge algorithms and artificial intelligence that produce racist or biased outputs, when provided with benign inputs.[12] Dr. Safiya Umoja Noble wrote on this subject in depth in her book, *Algorithms of Oppression*.[13] The book focuses heavily on the viewpoint of black women and girls experiencing racist outputs on Google Search, based on neutral inquiries[14]; such as the term, "black girl."[15] The search results brought up content

---

[6] *Id.*
[7] *Id.*
[8] *Id.* at 3.
[9] *Most Popular Social Media 2004/2022*, STATISTIC & DATA (2022), https://statisticsanddata.org/data/most-popular-social-media-in-history/ (stating, Facebook boasts a 2.9 billion user count, with YouTube, WhatsApp and Instagram following closely).
[10] Mosseri, *supra* note 1.
[11] Noble, *supra* note 5, at 1-229.
[12] *See* Noble, *supra* note 5, at 1- 229; Michael Kearns, et. al., *The Ethical Algorithm: The Science of Socially Aware Algorithm Design* (2020); Virginia Eubanks, *Automating Inequality: How High-Tech Tools Police, and Punish the Poor* (2018).
[13] *See* Noble, *supra* note 5, at 1- 229.
[14] *Id.* at 64-109.
[15] *Id.* at 66.

pages that were directly linked to long held biases and negative stereotypes about black women and girls, which Noble described in depth in the text.[16] This section will explore the arguments and theses in the *Algorithms of Oppression*[17] text and advance a critical analysis of the book.

A. Background and Synopsis

Algorithms of Oppression[18] was published in 2018[19] and was based largely on research Dr. Noble and her colleagues conducted in 2016 and prior.[20] Thus, much of the text centers on racism and the algorithms that comprise the Google Search engine.[21] The introductory chapters of the book explain why the inquiry into the intersectionality of racism and search algorithms is integral in the modern day.[22] Dr. Noble narrates her own experience attempting to search for positive content about black girls for her nieces, but instead was disappointed by the top-ranking search results which depicted black women in sexually explicit contexts.[23] Dr. Noble asserts that Google and other corporations responsible for algorithms producing racist effects may not actively be participating in overt racism at the programming stage.[24] Instead, these entities program their algorithms to support racist stereotypes and suggest discriminatory content because they directly benefit from the economic opportunities these harmful behaviors create.[25] Noble argues:

> At the core of my argument is the way in which Google biases search to its own economic interests—for its profitability and to bolster its market dominance at any expense… what is missing from the extant work on Google is an intersectional power analysis that accounts for the ways in which marginalized people are exponentially harmed by Google. Since I began writing this book, Google's parent company, Alphabet, has expanded its power into drone technology, military-grade robotics, fiber networks, and behavioral surveillance technologies such as Nest and Google Glass. These are just several of many entry points to thinking about the implications of artificial intelligence as a human rights issue.[26]

Further, Noble states that although many large companies, like Google, make important decisions to primarily further their economic interests, discriminatory hiring practices in the technology industry remain prevalent.[27] This

---

[16] *Id.* at 66-109.
[17] Noble, *supra* note 5, at 1-229.
[18] *Id.*
[19] *Id.*
[20] *Id.* at x.
[21] *Id.* at 64-109.
[22] *Id.*
[23] *Id.*
[24] *Id.*
[25] *Id.*
[26] *Id.* at 27.
[27] *Id.*

lack of diverse programmers affects the range of approaches employees will use to create algorithms, due to conscious and subconscious biases.[28] Noble supports this theory by articulating that, "Deep machine learning, which is using algorithms to replicate human thinking, is predicated on specific values from specific kinds of people—namely, the most powerful institutions in society and those who control them."[29]

Because these mammoth corporations use their technologies, like Google Search, as their primary mode of interaction with users, the algorithmic automated responses users receive are often unfiltered and unchecked.[30] Noble argues these corporations' loftiness and public inaccessibility pose significant issues when their algorithms are left to create discriminatory effects with no singular entity held accountable.[31]

In the wider discussion of oppressive algorithms, *Algorithms of Oppression*[32] has been cited 5,084 times by academic, business, and technology sources. Some of the most prominent citing sources include the text, *Race After Technology: Abolitionist Tools for the New Jim Code*[33], by Ruha Benjamin, and *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*[34], by Nick Couldry and Ulises Mejias. Both these sources are academic facing books, crafted to inform their audiences about the ill-effects of technology on human life, due to technological autonomy.[35] Further, many of these citing texts discuss a wide range of technological implications, including those relating to machines, cars, search engines, and algorithms. Very few of the sources discuss social media algorithms, and no one reference narrowly focuses on the deleterious effects of social media algorithms. Finally, *Algorithms of Oppression*[36] has not been cited by any legal writings.

B. Groundbreaking Elements in the Text

This section will assess the trailblazing, strong components in the *Algorithms of Oppression*[37] text.

1. Technological Interface from Black Perspectives

Dr. Noble approaches the analysis of oppressive algorithms and technologies through a novel lens, by narrating the negative experiences of black users entering search terms that mention blackness, into racially biased search

---

[28] *Id.*
[29] *Id.* at 29.
[30] *Id.*
[31] *Id.*
[32] *Id.* at 1-229.
[33] RUJA BENJAMIN, RACE AFTER TECHNOLOGY: ABOLITIONIST TOOLS FOR THE NEW JIM CODE 1 (2019).
[34] NICK COULDRY & ULISES MEJIAS, THE COSTS OF CONNECTION: HOW DATA IS COLONIZING HUMAN LIFE AND APPROPRIATING IT FOR CAPITALISM 1 (2019).
[35] *See generally*, Benjamin, *supra*, at 1; Couldry, *supra*, at 1.
[36] Noble, *supra* note 5, at 1-229.
[37] *Id.*

algorithms.[38] Noble's research and assertions about this unfortunate phenomenon are concentrated in the third chapter of *Algorithms of Oppression*[39], entitled, "Searching for People and Communities."[40] Numerous works[41] discussing biased algorithms examine user experiences, based on the results of certain search parameters from a white, western perspective. These inquiries are flawed, especially in the realm of the corporate technology industry, which purports to cater to and reach global audiences.[42] These sources formulate catchall queries based in assumptions that Eurocentric and white values are a centered normal, and all other perspectives are deviations from the center.[43] In contrast, Noble's text directly states that its focus centers on the overwhelmingly negative experiences of black users who interface with algorithmic technology and interrogates the cause.[44]

In "Searching for People and Communities,"[45] Noble discusses Google Search's harmful algorithm that blindly promotes unverified, racist sources in exchange for advertisement fees.[46] Noble argues that this practice is at issue because of the tangible havoc these invisible algorithms wreak in the real lives of historically oppressed individuals.[47] When this racially hateful, baseless rhetoric is uniformly pushed to the top of search results for all users, users often digest this disinformation as indicative of reality and truth.[48] When black users only see negative reinforcement about their identities, Noble argues, their self-images suffer.[49] Worse, Noble points to the alarming search behaviors of the Emanual African Methodist Episcopal Church mass-shooter Dylann Roof.[50] Who, prior to the shooting, self reportedly searched "black on white crime"[51] on Google Search.[52] As elucidated in Noble's book, "[w]hat Roof found was information that confirmed a patently false notion that Black violence on White Americans is an American crisis."[53] The Counsel of Conservative Citizens (CCC) disseminated the information on the site Roof visited.[54] The Southern Poverty Law Center describes the CCC as a "crudely white supremacist group."[55]

---

[38] *Id.* at 110-19.
[39] Noble, *supra* note 5, at 1-229.
[40] *Id.* at 110-19.
[41] *See* VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS POLICE, AND PUNISH THE POOR (2018); CAROLINE CRIADO PEREZ, INVISIBLE WOMEN: DATA BIAS IN A WORLD DESIGNED FOR MEN (2019); and others.
[42] *Id.*
[43] *Id.*
[44] Noble, *supra* note 5, at 1-229.
[45] *Id.* at 110-19.
[46] *Id.* at 115-19.
[47] *Id.*
[48] *Id.*
[49] *Id.*
[50] *Id.* at 110.
[51] *Id.*
[52] *Id.*
[53] *Id.*
[54] *Id.* at 112.
[55] *Id.* at 112 (stating, "Created in 1985 from the mailing lists of its predecessor organization, the CCC, which initially tried to project a "mainstream" image, has evolved into a crudely white

From the perspective of a black Google Search user, Noble argues, Google is indirectly taking place in the deleterious promotion of racist, emotionally polarizing information that has been linked to hate crimes against other black people.[56] Noble appropriately points to technology companies, like Google, that create algorithms which direct information to millions of people, as taking the place of traditional information sources, such as "teachers, books, history, and experience."[57] These traditional information sources, including teachers, schools, and libraries, assumed a form of liability to the public and were subjected to governmental checks.[58] Modern for-profit technology companies, like Google, do not assume the same kind of liability or oversight.[59]

Historically marginalized users have experienced an uncensored and insidious reversion in information sources available to the public, as the uncharted domain of the internet remains ungoverned.[60] Where black and other oppressed users could find solace in governmental regulations of public-school curriculums and other information sources, which oversaw the erasure of racist values in widely accessible databases, privately owned global information companies no longer must adhere to the same inclusive objectives.[61] Internet sources, such as Google Search, have eclipsed public resources, such as the library in popularity.[62] Noble justifiably argues:

> What we need is a way to reframe, reimagine, relearn, and remember the struggle for racial and social justice and to see how information online in ranking systems can also impact behavior and thinking offline. There is no federal, state, or local regulation of the psychological impact of the Internet, yet big-data analytics and algorithms derived from it hold so much power in overdetermining decisions. Algorithms that rank and prioritize for profits compromise our ability to engage with complicated ideas. There is no counterposition, nor is there a disclaimer or framework for contextualizing what we get.[63]

Overall, Dr. Noble has set forth a palpable issue that black users face in interacting with algorithms operated by for-profit corporations. The intersection between invisible algorithms that make fast and final choices on behalf of large, inaccessible corporations with deleterious effects on marginalized groups is an

---

supremacist group whose website has run pictures comparing the late pop singer Michael Jackson to an ape and referred to black people as "a retrograde species of humanity.")

[56] *Id.* at 115.

[57] *Id.*

[58] *Id.* at 115-16.

[59] *Id.*

[60] *Id.* at 117.

[61] *Id.*

[62] *Id.*

[63] *Id.*

issue that should be explored further than the contours of the *Algorithms of Oppression*[64] inquiry.

2. *Immunity of Information Technology Corporations*

An extension to the arguments advanced in the "Searching for People and Communities"[65] chapter appears in the subsequent chapter, "Searching for Protections from Search Engines."[66] Dr. Noble expands on the argument that information technology companies passively accept fees paid by any content creator for higher result rankings on search engines.[67] Instead, these corporations' advertisement profit models are closely linked with the active promotion of content types that draw the most attention, through scandalous or salacious themes.[68] As Noble elaborates:

> Yet, as I have noted in previous chapters, the prioritization and circulation of misrepresentative and even derogatory information about people who are oppressed and maligned in the larger body politic of a nation, as are African Americans, Native Americans, Latinos, and other peoples of color, is an incredible site of profit for media platforms, including Google. We need to think about delisting or even deprioritizing particular types of representative records.[69]

Noble advances a solution to delist or push racially demeaning information further down on the search results list, thus, stemming the negative impacts of the webpages by obscuring the disinformation from the view of most users.[70] However, Noble suggests this action in a long-term context, because many large corporations construct their feedback and communication channels to be one-sided.[71] As Noble presents, many users have attempted to report postings or provide feedback to Google Search, for example, to no avail. Noble deftly underscores the level of unbridled power these privately owned information companies wield in a global social context.[72] Not only are these companies not required to transparently report how their algorithms make decisions on which webpages to promote in search results, or directly answer takedown requests from individual users, but these entities also often escape legal liability for indirectly promoting and encouraging hateful speech that inspires hate crimes committed against human beings.[73] Even further, these corporate entities unrestrictedly control how users consume digital

---

[64] Noble, *supra* note 5, at 1-229.
[65] *Id.* at 110-19.
[66] *Id.* at 110-34.
[67] *Id.*
[68] *Id.*
[69] *Id.* at 124.
[70] *Id.* at 124-34.
[71] *Id.*
[72] *Id.*
[73] *Id.*

information and influence individuals' formulations of ideas and identities with which they approach the world.[74] Noble supports this notion, insisting:

> The questions about who controls the records of our social life and how they can be forgotten must move to the fore in the United States. They are explicitly tied to who can own identity markers and how we can reclaim them at both the individual and community level… We cannot ignore the long-term consequences… Ultimately, what I am calling for is increased regulation that is undergirded by research that shows the harmful effects of deep machine-learning algorithms, or artificial intelligence, on society.[75]

Dr. Noble has carved out a sound argument concerning the scale of reach privately owned information technology corporations possess on their global platforms; the unrestrained power they use in pushing content, ideas, and information to global users; and their lack of liability for the users they profit from when negative online propaganda can be directly connected to harmful acts committed by users. These deleterious corporate activities are not confined to search engine algorithm platforms but occur on other platforms not discussed in *Algorithms of Oppression*[76]. Noble's examination must be further expanded to meet contemporary needs.

C. *Points of Critique within the Text*

This section will analyze the discussions in the *Algorithms of Oppression*[77] text that should be expanded and focused on in future scholarship on the topic.

1. *Narrow Focus on the Negative Experiences of Black Girls and Women*

Dr. Noble prefaces that the research reported in *Algorithms of Oppression*[78] would center on black women and girls' experiences by describing her desire to find uplifting content on the Google Search platform for her black nieces.[79] Although racial biases and stereotypes negatively affecting black women and girls pervade the technology and artificial intelligence industries, these antagonistic anti-black woman values are inextricably linked to wider, malicious anti-black attitudes detrimentally impacting black men and boys.[80] Although Noble's work cursorily mentions the unfortunate programming choices that create

---

[74] *Id.*
[75] *Id.* at 130.
[76] Noble, *supra* note 5, at 1-229.
[77] *Id.*
[78] *Id.*
[79] *Id.* at 3.
[80] *See* Nation, *When Black Death Goes Viral, it can Trigger PTSD-like Trauma,* PBS (July 22, 2016, 8:04 PM), https://www.pbs.org/newshour/nation/black-pain-gone-viral-racism-graphic-videos-can-create-ptsd-like-trauma.

algorithms which produce racially biased search results affecting black men[81], a more detailed analysis on these discriminate activities is essential to depict the severity of the phenomenon.

As previously mentioned, several chapters in *Algorithms of Oppression*[82] directly underscore the linkage between the top-rated search results promoting racist disinformation and violent hate crimes committed by users who consume the rhetoric.[83] Yet, there is no discussion about why seemingly benign search parameters yield results that are connected to violence involving black individuals. Normalized stereotypes and ideologies still exist, that equate black maleness to fear-inducing violence committed against whites, which must be punished with even harsher violence.[84] Such stereotypes were the undercurrent of the information Dylann Roof read on Google Search promoted sites before he committed a mass shooting that killed black men and women.[85] In quotidian contexts, however, this connection of violence and black male identity reoccurs on a less glaring scale.

According to a viral experiment concerning an artificial intelligence service offered by Google Vision Cloud, the algorithm "labeled an image of a dark-skinned [seemingly male] individual holding a thermometer 'gun' while a similar image with a light-skinned individual was labeled 'electronic device.'"[86] This same mislabeling that implicates negative stereotypes of black men as approximate to violence was repeated in several other tests conducted in the Google Vision Cloud experiment.[87] Further, on review of the programming techniques used in computer recognition software, like Google Vision Cloud, the process involves humans labelling millions of items in pictures.[88] These human values and choices serve as the accuracy standards that computers use to reproduce the labelling in training data sets.[89] The training data sets often also contain skews and biases against non-white individuals such that, "[dark-skinned people are] featured much more often in scenes depicting violence in the training data set, [thus] a computer making

---

[81] Noble, *supra* note 5, at 11.
[82] Noble, *supra* note 5, at 1-229.
[83] *Id.* at 111-18, 133.
[84] *See Family Members of Buffalo Mass Shooting Want Focus on Preventing Racial Violence, Not Death Penalty,* DEATH PENALTY INFO (Sep. 07, 2022), https://deathpenaltyinfo.org/news/family-members-of-buffalo-mass-shooting-want-focus-on-preventing-racial-violence-not-death-penalty (stating, "Federal prosecutors allege that the accused 19-year-old shooter, Payton Gendron, who kept a detailed 700-page on-line diary of his plans, was motivated by a right-wing racist conspiracy theory that forces in the U.S. were attempting to replace white citizens with people of color and wanted to 'inspire others to commit similar attacks.' ABC reported that a Joint Intelligence Bulletin obtained from federal counterterrorism authorities raised fears that the contents of Gendron's diary, if publicly released, 'outlining [his] tactics, techniques and procedures … will likely enhance the capabilities of potential mass casualty shooters who may be inspired by this attack.'").
[85] Noble, *supra* note 5, at 115.
[86] Nicolas Kayser-Bril, *Google apologizes after its Vision AI produced racist results,* ALGORITHM WATCH (Apr. 07. 2020), https://algorithmwatch.org/en/google-vision-racism/#:~:text=In%20an%20experiment%20that%20became,was%20labeled%20%E2%80%9Celectronic%20device%E2%80%9D.
[87] *Id.*
[88] *Id.*
[89] *Id.*

automated inferences on an image of a dark-skinned hand is much more likely to label it with a term from the lexical field of violence."[90]

Algorithms of Oppression[91] failed to fully engage the issue of biased programing approaches which slant to include the view that blackness and the black male identity should be consistently associated with violence. This flawed human and artificial thinking influences results on search platforms and other forms of online artificial intelligence programs.[92]

2. Stunted Inquiry into Solely Search Programs

The *Algorithms of Oppression*[93] text narrowly focused on the biased nature of algorithms existing in search platforms.[94] The technology industry is one marked by its rapidly changing innovations and services.[95] Yet, Dr. Noble failed to anticipate new kinds of technology platforms that depend on coding and algorithmic decision-making, which could potentially usurp the older search technology.

According to a recent personal search of the terms, "unprofessional hairstyles" and "black men," on the Google Images platform, the top-rated results have been largely expunged and rectified from their previous biases.[96] Based on Noble's Google Images research of "unprofessional hairstyles" prior to 2018, the results overwhelmingly suggested depictions of black individuals with traditionally African American hairstyles, such as braids.[97] Google has retroactively begun fixing its algorithms to disconnect certain search terms from racist and biased results.[98] However, *Algorithms of Oppression*[99] fails to speak to current issues which arise when other technologies, such as social media platforms, have not received the backlash Google has to improve their algorithms, and continue to implement oppressive algorithms. These social media applications are particularly important because they have surpassed search engines, like Google Search, in popularity for sources of news, entertainment, and interaction.[100]

Based on current search parameters on Google Search that include terms like, "police brutality" and "police shootings," the top results spring up from webpages which are widely recognized as reputable.[101] These webpages chiefly are

---

[90] *Id.*
[91] Noble, *supra* note 5, at 1-229.
[92] Kayser-Bril, *supra* note 78.
[93] Noble, *supra* note 5, at 1-229.
[94] *Id.*
[95] Mosseri, *supra* note 1.
[96] Sikudhani Foster-McCray, *Current State of Google AI Searches*, (Feb. 28, 2023, 1:13 PM) (on file with author).
[97] Noble, *supra* note 5, at 81-83.
[98] Foster-McCray, *supra* note 89.
[99] Noble, *supra* note 5, at 1-229.
[100] Jhinuk Sen, *Gen Z Prefers Using TikTok, Instagram for search instead of Google, as per Google's own data,* BUSINESS TODAY (July 14, 2022, 4:58 PM), https://www.businesstoday.in/technology/news/story/gen-z-prefers-using-tiktok-instagram-for-search-instead-of-google-as-per-googles-own-data-341536-2022-07-14.
[101] Foster-McCray, *supra* note 89.

attached to web addresses with secure sources, such as ".orgs" and ".edus."[102] Contrarily, social media platforms, such as Facebook, Twitter, Instagram, and TikTok, allow source sharing without the need for separate web addresses.[103] Instead, information sharing on these service programs is formulated to be anonymous, instantaneous, and viral.[104] Upon entering the term, "police brutality," into the search feature on Instagram, the results simply display pictures, with no easily visible source attached, to check reputability.[105] Such as with privately-owned search algorithms, social media algorithms are programmed to push selected content to the top of the user interface view.[106] Even worse than biased search algorithms which suggest anti-black content, social media algorithms provide no warning for users as to the source or nature of the content they are pushed before they click.[107]

According to an article entitled, "A Deep Learning-Based Approach for Inappropriate Content Detection and Classification of YouTube Videos," authors, Kanwal Yousef and Tabassam Nawaz assert that:

> The exponential growth of videos on YouTube has attracted billions of viewers among which the majority belongs to a young demographic. Malicious uploaders also find this platform as an opportunity to spread upsetting visual content, such as using animated cartoon videos to share inappropriate content with children. Therefore, an automatic real-time video content filtering mechanism is highly suggested to be integrated into social media platforms.[108]

As indicated in the mentioned article, there is an important discussion taking place surrounding the implications of social media corporations that configure their platforms to support anonymity, virality, and covertness of uncensored content. This practice is alarming when applied to biased algorithms that may suggest negative content to unwitting users.

Unchecked, ungoverned, and unlabeled social media posts generated by millions of random users, pose just as significant a problem as the troubling search results provided on Google Search. The disinformation on social media is largely uncensored and consistently pushed to users without user prompting.[109] These methods of content pushing are often based on algorithmic decision-making, heavily rooted in traditional negative stereotypes and biases.[110] Dr. Noble failed to

---

[102] *Id.*

[103] *Compare*, Mosseri, *supra* note 1; *Our approach to recommendations*, *with* TWITTER HELP CENTER, https://help.twitter.com/en/rules-and-policies/recommendations; *and Video Discovery Tips*, YOUTUBE HELP, https://support.google.com/youtube/answer/11914225?hl=en.

[104] Mosseri, *supra* note 1.

[105] Sikudhani Foster-McCray, *Current State of Instagram AI Searches*, (Feb. 28, 2023, 1:15 PM) (on file with author).

[106] Mosseri, *supra* note 1.

[107] *Id.*

[108] Kanwal Yousef & Tabassam Nawaz, *A Deep-Learning Based Approach for Inappropriate Content Detection and Classification of YouTube Videos*, 10 IEEE ACCESS 16283, 16298 (2022).

[109] Mosseri, *supra* note 1.

[110] Kayser-Bril, *supra* note 78.

mention or assess this contemporary challenge in *Algorithms of Oppression*[111]. Thus, this Paper will address that imminent issue.

II. THESIS: SOCIAL MEDIA ALGORITHMS RECOMMEND NEGATIVE AND HARMFUL CONTENT TO USERS, ROOTED IN RACIAL STEREOTYPING AND VIRALITY

Search platforms have been linked to discriminatory algorithms that negatively impact user experience.[112] As previously explored, the detrimental racial biases are embedded into the programming of algorithms, due to manmade choices.[113] An additional issue occurs when social media platforms are formulated to increase global engagement by suggesting unprompted content to users that is unfiltered, inflammatory, and racially degrading. Where search platforms order links on a page from top to bottom and give the user a choice which source of information to consume, social media platforms do not provide users a choice before a viral post is displayed on an explore page or reels feature.[114] Further, popular posts are actively pushed by an algorithm to thousands of users at a given time, without the action of the user.[115] This practice involves the actions of the algorithm much more than those of the user.[116] This section will discuss the derogatory content that is specific to social media platforms and analyze how the algorithms work to reinforce black racial stereotypes for the benefit of global social media platforms.

A. Harmful Racial Stereotype Reinforcement through Virality

Many contemporary black stereotypes are based in overrepresented tragedies.[117] According to a 2015 article entitled, "Guns and race: The different worlds of black and white Americans," black Americans face a higher proportional mortality rate due to homicide than white Americans.[118] This statistic is grounded in research data provided by the Centers for Disease Control's (CDC) National Violent Death Reporting System. In 2020, of the 16,108 reported deaths of black Americans, 11,579 were a result of homicide.[119] Conversely, only 7,071 homicides were reported of the 43,313 total white American death count during 2020.[120] The disparity between white and black American mortality causes is stark, however

---

[111] Noble, *supra* note 5, at 1-229.
[112] *Id.*
[113] Mosseri, *supra* note 1.
[114] *Id.*
[115] *Id.*
[116] *Id.*
[117] Richard V. Reeves & Sara E. Holmes, *Guns and race: The different worlds of black and white Americans*, BROOKINGS (Dec. 15, 2015), https://www.brookings.edu/blog/social-mobility-memos/2015/12/15/guns-and-race-the-different-worlds-of-black-and-white-americans/.
[118] *Id.*
[119] *WISQARS National Violent Death Reporting System: Violent Deaths Report,* CTRS. FOR DISEASE CONTROL AND PREVENTION, https://wisqars.cdc.gov/nvdrs/?rt=3&rt2=0&y=2020&g=00&i=0&m=20810&s=0&r=2&e=0&rl=0&pc=0&pr=0&h=0&ml=0&a=ALL&a1=0&a2=199&g1=0&g2=199&r1=NVDRS-INTENT&r2=NONE&r3=NONE&r4=NONE.
[120] *Id.*

when stereotypes spring from incidents, without context, they become harmful disinformation. Based on the same CDC reporting, the 11,579 homicides only constituted 0.02847% of the black American population in 2020, standing at 40,670,272.[121]

Stereotypes about increased violence in the black community are not new.[122] These same misconceptions about an entire group of individuals are re-entrenched through information sources and media.[123] Although there is some truth to higher homicide rates in select black Americans, these rates are irreflective of the wider populations' experiences and identities.[124] The most salacious, violent, and intriguing stories are often used as entertaining narratives of the larger whole.[125] With respect to social media corporations, high post engagement produces higher revenues.[126] Because content steeped in stereotypes, violence, and sexuality produces higher rates of clicks, reposts, and comments, social media algorithms are programmed to push it to users.[127]

According to Instagram's self-published 2021 article, "[Instagram's algorithm for] Reels is designed to entertain you. Much like Explore, the majority of what you see is from accounts you don't follow. So we go through a very similar process where we first source reels we think you might like, and then order them based on how interesting we think they are to you."[128] Instagram has explained its algorithms' purpose for entertainment and its configuration based on "ordering"[129] content on predicted user interest.[130] This ordering is similar to the ranking functions of the search algorithms mentioned in the *Algorithms of Oppression*[131] text. As Dr. Noble explained in *Algorithms of Oppression*[132], concerning search engines' business models based in advertising rather than reputable information sharing:

> Yet much of the content surfaced in a web search in a commercial search engine is linked to paid advertising, which in part helps drive it to the top of the page rank, and searchers are not typically clear

---

[121] *Id.*

[122] *See* Jennifer Bloomquist, *The Minstrel Legacy: African American English and the Historical Construction of "Black" Identities in Entertainment*, AFRICANA STUD. FAC. PUBS., 1, 25 (2015).

[123] *Id.*

[124] *WISQARS, supra* note 111.

[125] Mosseri, *supra* note 1.

[126] *Id.*

[127] *See* Peter Dizikes, *Why social media has changed the world- and how to fix it (MIT professor Sinan Aral's new book "The Hype Machine," explores the perils and promise of social media in a time of discord.*, MIT NEWS OFF. (Sep. 24, 2020), https://news.mit.edu/2020/hype-machine-book-aral-0924 (discussing, "'This is a well-designed, well-thought-out machine that has objectives it maximizes,' Aral says. 'The business models that run the social-media industrial complex have a lot to do with the outcomes we're seeing — it's an attention economy, and businesses want you engaged. How do they get engagement? Well, they give you little dopamine hits, and … get you riled up. That's why I call it the hype machine. We know strong emotions get us engaged, so [that favors] anger and salacious content.'").

[128] *Id.*

[129] *Id.*

[130] *Id.*

[131] Noble, *supra* note 5, at 1-229.

[132] *Id.*

> about the distinctions between "real" information and advertising. Given that advertising is a fundamental part of commercial search, using content analysis to make sense of what actually is served up in search is appropriate and consistent with the articulation of feminist critiques of the images of women in print advertising. These scholars have shown the problematic ways that women have been represented—as sex objects, incompetent, dependent on men, or underrepresented in the workforce—and the content and representation of women and girls in search engines is consistent with the kinds of problematic and biased ideas that live in other advertising channels. Of course, this makes sense, because Google Search is in fact an advertising platform, not intended to solely serve as a public information resource in the way that, say, a library might. Google creates advertising algorithms, not information algorithms.[133]

Social media platforms use similar profit models and approaches in configuring user interface and algorithms to maximize advertising revenues.[134] Thus, the content that is predicted to procure the most interactions or advertising information for third-party companies, will be recommended to users more heavily than less ostentatious content. According to an article published by data researchers, Furkan Puligu and Cihan Varol, social media platforms do not simply rank content based on its popularity to then attach outside advertisement to the media, like with search engines.[135] Instead, social media platforms go a step further. They both track users' interface with content and store that information on their databases.[136] The article describes user interface and data storage on Instagram stating:

> On average, an individual living in the U.S. spends approximately two hours on social media sites every day, while leaving a substantial number of digital artifacts on various sites…Due to advantageous storage capability and fast access to persistent data with the utilization of JavaScript and JSON…One of the websites that recently joined this trend is the popular social media site Instagram, which has over 1074 billion users worldwide.[137]

Since social media platforms approach algorithmic programming with the support of their data gathering and advertisement capabilities at the core, there is less focus on what kinds of media are being promoted, even if the media constitutes misinformation or harmful content. As Dr. Noble stated in the quotation above concerning search engines, "Given that advertising is a fundamental part of commercial search, using content analysis to make sense of what actually is served up in search is appropriate and consistent with the articulation of feminist critiques

---

[133] *Id.* at 38.

[134] Furkan Paligu & Cihan Varol, *Browser Forensic Investigations of Instagram Utilizing IndexedDB Persistent Storage, Cybersecurity and Cybercrime in the Age of Social Media*, FUTURE INTERNET (2022), at 1.

[135] Paligu, et. al., *supra*, at 1.

[136] *Id.*

[137] *Id.*

of the images of women in print advertising."[138] Although Noble's text mainly focuses on the top-ranked sites with negative stereotypes of black women, this Paper focuses on the promotion of misinformation through stereotypes of both black female and male identities. As discussed above, current issues often serve as the front stage for political debates and ideology that can quickly divulge into racist or hateful rhetoric based in stereotypes.[139]

Although the proportional homicide rate in the black American community is higher than that of the white American community[140], current issues such as police brutality and gun violence involving black individuals, have over emphasized the reality of black American mortality in an alarming manner. This amplification of black death is rooted partly in the advancement of video capable handheld phones that can capture tragic incidents and support instant reposts to social media platforms.[141] The issue this Paper analyzes is how social media algorithms filter and push this violent and disturbing content for entertainment purposes, rather than for educational or teaching purposes. Further, this Paper interrogates which users this alarming content is pushed to, based on the algorithms' predictions on, as Instagram states, "how interesting we think [posts] are to you."[142]

Instagram has stated in its own article that its algorithm participates in resharing and amplifying content about "social unrest"[143] for entertainment and virality purposes. Part of Instagram's explanation stated:

> Stories that were "reshared" from Feed: until recently, we valued these Stories less, because we've heard consistently that people are more interested in seeing original Stories. But we see a swell of reshared posts in big moments – everything from the World Cup to social unrest – and in these moments people were expecting their Stories to reach more people than they did, so we stopped… Another important case to call out is misinformation. If you post something that third-party fact checkers label as misinformation, we don't take it down.[144]

Instagram, alongside other social media platforms, including Twitter, Facebook, and YouTube, allow or push some form of misinformation or disturbing content on their platforms.[145] It is important, as Dr. Noble stated, to identify the "problematic ways that [marginalized people] have been represented"[146] on social

---

[138] Noble, *supra* note 5, at 38.
[139] Mosseri, *supra* note 1.
[140] *WISQARS, supra* note 111.
[141] Cody Mello-Klein, *How do Videos of Police Brutality Affect Us, and how Should We Engage with Them?*, NE. GLOB. NEWS (Feb. 06, 2023), https://news.northeastern.edu/2023/02/06/police-brutality-videos-impact/ (discussing videos of police brutality and stating, "footage of a young Nichols skateboarding went viral and offered a glimpse of a man enjoying life, the video of the brutal beating lingers in the minds of many.")
[142] Mosseri, *supra* note 1.
[143] *Id.*
[144] *Id.*
[145] *Id.*
[146] Noble, *supra* note 5, at 38.

media platforms in the pursuit of maximizing the profit models of these companies' algorithms.

B. How Social Media Algorithms Perform Actions and Push Negative Content

This section will analyze instances in which tragic or racially degrading content was pushed to users for entertainment purposes and the impact the virality caused on black identities. It further will explain how algorithms perform this content recommendation function and how algorithmic programming is racially marginalizing at its core.

1. Historical Representative Instances of Racial Content

Like Dr. Noble's personal impetus for her inquiry into how search parameters have racially oppressive effects, this Paper's query arises from personal experiences and a desire for answers. Although racially driven police brutality has remained a pervasive issue in American history, the advent of smartphones with video capabilities and the share space of the internet have forced the discourse around the topic to acknowledge that the violence occurs.[147]

The murder of Trayvon Martin in 2012, was a hotly debated topic.[148] Although Martin's, death was not archived on video, the emergency call from Martin's killer, George Zimmerman, was circulated on websites and social media platforms.[149] At the height of the pending criminal case against Zimmerman, one of the popular search parameters entered into the Google Images search engine was, "Trayvon Martin."[150] Some of the top results returned images of the seventeen-year old's lifeless body lying on the corner where he was killed.[151] Martin's eyes were open.[152] The media's circulation of images depicting brutally murdered underaged black people is not a phenomenon that came to the fore with the introduction of the internet.[153] Instead, a famous instance was the circulation of Emmett Till's mangled

---

[147] *Black Lives Upended By Policing: The Raw Videos Sparking Outrage*, N.Y. TIMES (Apr. 19, 2018), https://www.nytimes.com/interactive/2017/08/19/us/police-videos-race.html (stating, "Raw videos that show officers shooting and beating unarmed black people have stirred outrage and prompted disbelief. Captured by cellphones or police cameras, footage has spread through social media, shining a light on disturbing police encounters."

[148] *ABC World News: Trayvon Martin Shooting Video: New Evidence* (ABC News television broadcast May 18, 2012).

[149] *Id.*

[150] Karen Grisby Bates, *A Look Back At Trayvon Martin's Death, and the Movement it Inspired*, NAT'L PUB. RADIO (July 31, 2018, 7:34 AM), https://www.npr.org/sections/codeswitch/2018/07/31/631897758/a-look-back-at-trayvon-martins-death-and-the-movement-it-inspired.

[151] Sikudhani Foster-McCray, *Trayvon Martin Search Results*, (Feb. 28, 2023, 1:13 PM) (on file with author).

[152] *Id.*

[153] *Id.*

face on the cover of *Jet* magazine, after his mother insisted on an open-casket funeral for her son, who was killed in 1955.[154]

Historical attitudes about police or vigilante violence against black individuals were rooted in denial.[155] However, *Jet* magazine was one of the only news sources, at the time, that was willing to take on the gruesome truth of the white mob lynching of a black child.[156] According to an article published by Time magazine:

> When Till's mother Mamie came to identify her son, she told the funeral director, "Let the people see what I've seen." She brought him home to Chicago and insisted on an open casket. Tens of thousands filed past Till's remains, but it was the publication of the searing image photographed by David Jackson and first published in Jet magazine, with a stoic Mamie gazing at her murdered child's ravaged body, that forced the world to reckon with the brutality of American racism.[157]

The punctuating difference between the media circulation of Till's horrific photographs and the disturbing videos of black death circulated by contemporary social media platforms is that Mamie Till authorized *Jet* to use her son's images for educational, social reform reasons.[158] Further, *Jet* had a financial interest in controlling the exploitation of the images.[159] In contrast, social media platforms have an opposite interest in encouraging the exploitation and dissemination of death content for an increase in virality and user interactions.[160] Where the families of victims could contract to constrict the scope of the media's usage of images or videos of a death, social media algorithms make separate, instantaneous decisions for recommendations that do not strictly adhere to families' wishes.[161]

Finally, a stark difference between the Trayvon Martin content ranked on Google Search and the content regarding the most contemporary killings on social media platforms is that the user must actively enter parameters that directly call up results containing death and killing.[162] As discussed above, social media platforms, like Instagram, feature functions that signal content for "social unrest"[163] and increase engagement by recommending content to users that are likely to interact with the content. This imbalance in the actions of the user versus the algorithm may inaccurately inflate negative stereotypes attached to the

---

[154] Time Photo, *When One Mother Defined America: The Photo That Changed the Civil Rights Movement*, TIME MAG. (2016) at 1.
[155] *Id.*
[156] *Id.*
[157] *Id.*
[158] *Id.*
[159] *Id.*
[160] Mosseri, *supra* note 1.
[161] *Id.*
[162] Foster-McCray, *supra* note 142.
[163] Mosseri, *supra* note 1.

individuals depicted in the content, because the narrative becomes ubiquitous on popular platforms.

2. *Contemporary Representative Instances of Racial Content*

Social media platforms with algorithms that continuously and automatically perform recommendation functions have maintained their popularity in current times due to enhanced interface features.[164] This Paper will analyze the most prolific racialized, disturbing content that was distributed on Facebook, Instagram, Twitter, and TikTok.

The Black Lives Matter movement invigorated discourse about police brutality and unjust killings of black individuals.[165] Social media platforms aided in the prominence of the video evidence depicting black deaths at the hands of armed officers.[166] Yet, as discussed with the Instagram programming methods for its recommendations algorithm, the content was flagged for its engagement, as "social unrest"[167] content, and pushed to users.[168] This jarring jolt in recommended police killing videos exponentiated the recirculation of recorded black death to users at large, despite whether they were interested in the grotesque details of the killings.[169]

The tragic murder of Philando Castile is a widely studied incident of black death made viral on social media platforms, including Facebook, Twitter, and Instagram.[170] According to a Public Broadcasting Station (PBS) article, "the shooting death of Philando Castile was streamed live by his girlfriend on Facebook. The video, which shows Castile gasping for air after being shot four times by a Minnesota police officer, has since been shared on Facebook more than 5 million times."[171] Although the social media coverage of Castile's last moments was actively pushed by recommendation algorithms at the height of the Black Lives Matter social unrest, conclusory information about justice for Castile did not get the same social media push or virality.[172] As compared previously, the shocking violence of black death has historically been used by media platforms to capture audiences.[173] Yet the lack of control social media platforms now possess over the

---

[164] *Id.*

[165] Kenya Downs, *When Black Death goes Viral, it can Trigger PTSD-like Trauma,* PUB. BROAD. STATION (July 22, 2016, 8:04 PM), https://www.pbs.org/newshour/nation/black-pain-gone-viral-racism-graphic-videos-can-create-ptsd-like-trauma.

[166] *Id.*

[167] Mosseri, *supra* note 1.

[168] *Id.*

[169] Downs, *supra* note 155.

[170] *Id.*

[171] *Id.* (stating, "When video of the Baton Rouge shooting death of Alton Sterling first surfaced on July 5, social media networks became immediately populated with Sterling's final moments. The following day, the shooting death of Philando Castile was streamed live by his girlfriend on Facebook. The video, which shows Castile gasping for air after being shot four times by a Minnesota police officer, has since been shared on Facebook more than 5 million times.")

[172] *Id.*

[173] Time Photo, *supra* note 154.

actions of their algorithms and the subsequent exploitations of users consuming the pushed content, create a palpable issue rooted in the negative portrayal of black identities. According to the PBS article:

> Escaping the imagery can be nearly impossible, especially as online users post commentary and news updates. For some, it can merely be a nuisance. But research suggests that for people of color, frequent exposure to the shootings of black people can have long-term mental health effects. According to Monnica Williams, clinical psychologist and director of the Center for Mental Health Disparities at the University of Louisville, graphic videos (which she calls vicarious trauma) combined with lived experiences of racism, can create severe psychological problems reminiscent of post-traumatic stress syndrome… During the week of Sterling's and Castile's deaths, a scroll through timelines of black social media users could uncover subtle expressions of mental and psychological anguish, from pleas for others not [to] share these videos, to declarations of a social media hiatus.[174]

Although these viral, traumatic videos surfaced in millions of users' recommended pages at the height of the aftermath of Castile's slaying in 2016, this eye-catching violent content remains the fodder for social media popularity algorithms.[175] More recently, George Floyd's 2020 slaying, and Tyre Nichols' 2023 death by beating, were caught on graphic video and circulated millions of times on Facebook, Twitter, and Instagram.[176] According to several news sources, the George Floyd and Tyre Nichols murders received a higher volume of coverage and circulation on social media platforms than the unrecorded police raid that killed Breonna Taylor in 2020.[177] This disproportionality in the level of virality of these issues of social unrest may be linked to the sensationalism of black death voyeurism, blindly marketed as entertainment by uncontrolled algorithms. This phenomenon of black death and suffering as a spectacle is rooted in centuries old, anti-black stereotypes.[178]

A further difference between the controlled dissemination of images of black death, such as with *Jet* magazine and Emmitt Till, and the uncontrolled automated pushing by social media algorithms, is that the code cannot preset how many recirculations of the content will reach audiences.[179] Historical news platforms heavily controlled how many and where reproductions of their works entered the market.[180] Social media algorithms instantly push reproductions of

---

[174] *Id.*

[175] Sara Morrison, *Questions to Ask Yourself Before Sharing Images of Police Brutality*, VOX (June 11, 2020, 9:00 PM), https://www.vox.com/recode/2020/6/11/21281028/before-sharing-images-police-brutality-protest-george-floyd-ahmaud-arbery-facebook-instagram-twitter.

[176] *Id.*

[177] *Id.*

[178] Bloomquist, *supra* at 114.

[179] Mosseri, *supra* note 1.

[180] Time Photo, *supra* note 145.

content millions of times to global users, without pretext.[181] In the context of black death videos, "[t]hese images appear and autoplay on millions of social media timelines and television screens, making them almost impossible to avoid."[182] The detrimental and racially oppressive effects arise when social media algorithms apathetically overproduce an uncontextualized negative subset of black life by which the identity becomes pervasive and synonymous with the entire racial community.

More directly, the algorithmic pushing of this graphic content has been directly connected with trauma development in black users.[183] According to a recent article discussing social media disinformation practices, "[r]epeated exposure to images of Floyd's death and other violent incidents, along with angry rhetoric about the protests that followed, sometimes left [black users] short of breath and wanting to cry."[184] The few current solutions to the unbridled, irresponsible actions of social media platforms' algorithms have placed the burden on users.[185] Solutions proposed by social media platforms, such as Facebook, Instagram, Twitter, and TikTok, require the user to take active steps to prevent, "the company's algorithms [from deciding] what [users] see first."[186] These solutions unfairly release social media companies from liability and do not prevent users from initial exposure to traumatic content, pushed by algorithms.

3. How Social Media Algorithms "Push"

Social media platform algorithms complete functions to fulfill many purposes, including advertisement promotion, user information collection, and social influence.[187] Social media platforms are especially unlike traditional internet search platforms or print media, because they collect, store, and actively manipulate information about their users' races, religions, and language patterns.[188] Recently, platforms like Facebook, have faced legal liability based on the actions of their algorithms that allow manipulated targeting or exclusion of advertising content to users based on their race, religion, and socioeconomic status.[189] These algorithms are capable of categorizing the content of media and the identities of users based on racial stratifications, to drive information based on these classifications.[190]

---

[181] Morrison, *supra* at 164.
[182] *Id.*
[183] Thomas Germain, *How To Filter Hate Speech, Hoaxes, and Violent Clips Out of Your Social Media Feeds*, CONSUMER R. (July 21, 2022), https://www.consumerreports.org/social-media/filter-hate-speech-hoaxes-violent-clips-out-of-social-media-a6432199534/.
[184] *Id.*
[185] *Id.*
[186] *Id.*
[187] Mosseri, *supra* note 1.
[188] Ifeoma Ajunwa, *The Paradox of Automation as Anti-Bias Intervention*, 41 CARDOZO L. REV. 1671 (2020).
[189] Anthony E. Gambino, *Dante's Digital Inferno: Content Moderators' Class Action Against Facebook Feeds from Hell*, ST. JOHN'S LAB. & EMP. L. FORUM (June 28, 2022), https://stjclelblog.org/2022/06/dantes-digital-inferno/.
[190] Ajunwa, *supra* at 1671-1742.

Further, social media algorithms are programmed to disseminate information in a manner that affects decision making in identified social communities.[191] This programming, with the intended effect of controlling the kind of information that is proliferated, and to whom, can create devastating impacts in the social communities most targeted.[192] Social media algorithms promote and recirculate content in high volumes, at strategic times with the objective to socially influence users.[193] These algorithmic actions are founded in statistical studies, "social networks are generally believed to impact decision-making… it was observed that individual [idea] adoption is much more likely when participants received social reinforcement."[194]

Secondarily, although social media algorithms function largely to maximize virality and user interface, the underlying action of identifying selected users and predicting their content preferences based on racial, religious, and socioeconomic factors creates oppressive and detrimental effects in the identified groups. Many social media platforms describe how their algorithms perform this prediction function that guides what content is recommended, and to who.[195] However, these explanations are intentionally vague about which factors guide consistent, repeated algorithmic choices. An article published by Twitter discussed its algorithm's predictive analysis, stating, "[o]ur recommendations are based upon a variety of signals, including, but not limited to, interests you choose during onboarding, accounts & Topics you follow, Tweets you've liked, retweeted, or otherwise engaged with, and content that is popular in your network."[196] Racial categorizations, among others, are integral to algorithmic predictions for pushed disturbing content that features issues, such as race.[197] According to a 2019 Temple University study concerning social media predictive models, "[u]sing online data is an appealing [method], but such text is typically noisy and biased."[198]

Finally, the algorithms are manmade programs that are based in code written by fallible programmers.[199] At the baseline of social media coding, programmers teach algorithms to make distinctions and decisions about users.[200] These distinctions are called "class labels,"[201] which are defined by human programmers, and used to teach programs how to make decisions.[202] According to a California Law Review article, "[t]hrough this necessarily subjective process of

---

[191] Soumajyouti Sarkar, et. al., *Impact of Social Influence on Adoption Behavior: An Online Controlled Experimental Evaluation*, OFF. OF SCI. AND TECH. I. 1 (2021).
[192] *Id.*
[193] *Id.*
[194] *Id.*
[195] *Our Approach To Recommendations, supra* at note 3.
[196] *Id.*
[197] *Id.*
[198] Marija Stanojevic, et. al., *Surveying Public Opinion Using Label Prediction on Social Media Data*, CTR. FOR DATA ANALYTICS AND BIOMED. INFORMATICS 1, 2 (2019).
[199] Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671 (2016).
[200] *Id.*
[201] *Id.* at 677.
[202] *Id.*

translation, data miners may unintentionally parse the problem in such a way that happens to systematically disadvantage protected classes."[203] Algorithms have been found to create disparate impacts in protected classes due to deeply entrenched negative racial profiling and stereotyping woven through millions of lines of invisible code.[204] Social media algorithms, which are often manipulated and created in the discriminatory manner described, could possibly expose their corporate owners to several kinds of legal liability.

III. LEGAL IMPLICATIONS OF DETRIMENTAL ACTIVITIES INVOLVING SOCIAL MEDIA ALGORITHMS

The documented instances of race-based algorithmic recommendations and decision-making have been discussed both in this Paper and by other scholars. The harms that arise directly from the algorithmic promotion of disinformation and recirculations of racialized violence must be discussed to assess possible legal consequences. This section will indicate harms associated with this algorithmic activity, discuss possible legal claims individuals and classes may levy against social media corporations, and describe the current landscape of immunity afforded to computer service providers.

A. Negligence Harms and Corporate Tortious Liability Based in Algorithmic Actions

Although algorithms, code, and social media content exist on computer databases, as discussed, the disturbing content available on social media platforms depict real-time, trauma-inducing acts of human beings.[205] Because detached viewership of recorded deaths and violence do not pose a direct physical threat to most viewers, those primarily psychological harms experienced, must fall under limited forms of tortious negligence liability.[206] Further, although algorithms perform the recommendation or pushing functions for content, they are not legal entities and cannot be held liable like their corporate owners.[207] Thus, affected viewers may only file non-frivolous claims against social media corporations, rooted in negligent psychological harms suffered from a breach of corporate duty.[208] This legal liability is especially limited, in the field of social media, because platforms reach millions of global users.

Possible psychological harms that viewers may develop include post-traumatic stress disorder, acquired from un-consentingly experiencing recorded

---

[203] *Id.* at 678.
[204] *Id.*
[205] Downs, *supra* note 155.
[206] *See* Cf. Sullivan v. Boston Gas Co., 605 N.E.2d 805 (Mass. 1993).
[207] *See* Rebecca Crootof, *War Torts: Accountability for Autonomous Weapons*, 164 U. PA. L. REV. 1347 (2016).
[208] *See* Cf. Sullivan v. Boston Gas Co., 605 N.E.2d 805.

deaths or violence.[209] In connection with reoccurring depictions of racialized violence, a George Washington University professor of counseling urged, "[p]articularly for Black individuals…watching videos of police brutality can be traumatizing."[210] However, in a legal claim context, this trauma must be scientifically substantiated and marked with injurious symptoms.[211]

In a recent claim against Facebook, the plaintiff, Selena Scola and a putative class of similarly situated individuals, argued they developed documented post-traumatic stress disorder diagnoses from repeatedly watching videos depicting killings.[212] This putative class were all contracted moderators that were exposed to the videos in the course of their employment, and developed symptoms, such as loss of sleep, frequent irritability, and an inability to concentrate or finish work tasks.[213] Facebook reached a settlement with Scola and the listed plaintiffs, totaling $52 million for negligent exposure practices.[214] The merits of the claim hinged on the plaintiffs' statuses as individuals experiencing a work-related harm that resulted in professional trauma diagnoses.[215]

According to licensed psychologist, Dr. Ifetayo Ojelade, for an individual to be properly diagnosed with an extensive trauma, such as post-traumatic stress disorder (PTSD), the individual must meet the criteria set forth in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5).[216] For PTSD, the DSM-5 criteria, sets forth that repeated exposure to details of a traumatic event applies, except where, "[the] exposure [is] through electronic media, television, movies, or pictures, unless this exposure is work related."[217] Thus, in the context of recirculated violent videos pushed to viewers by algorithms, work-related exposures can include work as a content moderator or even as a content creator.[218]

In the context of this Paper, a likely negligence claim is based in negligent infliction of emotional distress for black moderators or content creators who develop PTSD from repeated exposure to racially charged violent videos pushed directly to them, based on class labels.[219] Like the facts in Scola's claim against Facebook, plaintiffs need not be direct employees of the defendant corporation to

---

[209] Amanda Macmillian, *A Facebook Content Moderator Says Her Job Gave Her PTSD. Here's What Experts Think*, HEALTH (Sep. 27, 2018), https://www.health.com/condition/ptsd/facebook-moderator-ptsd-lawsuit.

[210] Brook Endale, *What to ConsiderBefore Watching Videos of Police Brutality*, GEO. WASH. UNIV. (Feb. 14, 2023), https://phys.org/news/2023-02-videos-police-brutality.html.

[211] Macmilian, *supra* note 195.

[212] Selena Scola, et al. v. Facebook, Inc., Cal. Super. Ct., Cty. of San Mateo Case No. 18-civ-05135.

[213] Casey Newton, *Facebook will Pay $52 Million in Settlement with Moderators who Developed PTSD on the Job*, THE VERGE (May 12, 2020, 3:39 PM), https://www.theverge.com/2020/5/12/21255870/facebook-content-moderator-settlement-scola-ptsd-mental-health.

[214] *Id.*

[215] Selena Scola, et al. v. Facebook, Inc., *supra* note 212.

[216] Audio tape: Interview with Dr. Ifetayo Ojelade from A Healing Paradigm Psychology, Concerning Black PTSD (Mar. 31, 2023) (on file with author).

[217] Deion S. Hawkins, *"After Philando, I Had to Take a Sick Day to Recover": Psychological Distress, Trauma and Police Brutality in the Black Community*, NAT'L. LIBR. OF MED. 1113, 1127 (2021).

[218] *Id.*

[219] *See Cf. Sullivan v. Boston Gas Co.*, 605 N.E.2d 805, *supra* note 206.

show a harm and duty exists.[220] Further, in special circumstance cases, such as in Scola's claim, the plaintiff does not have to show a risk-laden physical event occurred that caused the plaintiff's emotional distress.[221] Finally, the claim would not be subject to immunity defenses currently afforded to internet service providers because, moderators and content creators are uniquely exposed to manipulated videos and flagged content which are developed by social media companies for moderator and content creator engagement.[222] Like with Scola, content moderators and creators are specifically engaged by social media companies and sent files of engineered content compiled by the companies' algorithms for moderator sorting or creator promotion.[223] Black moderators and creators are especially at risk because of the racially-based pushing and traumatic content specific to black individuals.

B. Criminal Harms and Corporate Civil Immunity Based in Algorithmic Actions

On a larger scale than individual psychological harms, racially motivated mass shootings have been linked to online oppressive disinformation, recommended to extremists through algorithmic prioritizations.[224] As Dr. Noble discussed in *Algorithms of Oppression*[225], Dylann Roof's decision to commit the church shooting was heavily influenced by, "allegedly typ[ing] 'black on White crime' in a Google search to make sense of the news reporting on Trayvon Martin."[226] The recommended information Roof found "confirmed a patently false notion that Black violence on White Americans is an American crisis."[227] More recently, Payton Gendron conducted a race-based shooting at a Tops Friendly Supermarket in 2022, that killed 10 black people.[228] Like Roof, Gendron kept an online diary that documented his dissent into hatred and referenced extremist groups and ideologies that were recommended to him on social media.[229] In the aftermath of the shooting, a family member of one of the victims described the 18-year-old shooter as "brainwashed"[230] because he was repeatedly fed an online "racist conspiracy theory that forces in the U.S. were attempting to replace white citizens with people of color."[231]

---

[220] *Id.*
[221] *Id.*
[222] Newton, *supra* note 199.
[223] *Id.*
[224] Noble, *supra* note 5, at 110.
[225] *Id.*
[226] *Id.* at 111.
[227] *Id.*
[228] The Associated Press, *Buffalo Shooting Suspect says his Motive was to Prevent 'Eliminating the White Race'*, NAT'L PUB. RADIO (June 16, 2022, 11:05 PM), https://www.npr.org/2022/06/16/1105776617/buffalo-shooting-suspect-says-his-motive-was-to-prevent-eliminating-the-white-ra.
[229] *Family Members of Buffalo Mass Shooting Want Focus on Preventing Racial Violence, Not Death Penalty, supra* note 77.
[230] *Id.*
[231] *Id.*

Although Gendron is currently facing a slew of federal hate crime charges, the victims, families of victims, and general public have questioned which other parties may be held liable for these criminal harms.[232] Similar cases have developed, asserting vicarious legal liability against social media corporations whose algorithms actively recommend and promote injurious disinformation, directly connected with resulting violent hate crimes.[233] *Gonzalez v. Google, LLC*, 2 F.4th 871 (9th Cir., 2021) is a case in which the plaintiff, an immediate family member of the deceased victim, sought to hold defendant corporations, Google, Twitter, and Facebook, directly and secondarily liable under the Anti-Terrorism Act.[234] The claims are based in allegations that Defendant Google, LLC, used its algorithms to recommend content to ISIS users, who then committed a planned terrorist attack that killed American citizens.[235]

Because the defendant corporations qualified as internet service providers, the defendants argued that 47 U.S.C §230 applied for preclusion from direct or secondary liability because they were arguably not "publishers"[236] or "speakers"[237] for the content.[238] The holding indicated that §230 struck the plaintiffs' claims against the defendants where there was no clear showing of economic gain in some claims and the evidence only showed that content-neutral algorithms acted similarly with all content, rather than to specifically support ISIS.[239] Although the plaintiffs were able to present that Google "shared advertising revenue with ISIS,"[240] the plaintiffs could not show a deeper, ISIS specific motivation for the algorithms' actions, to support direct or secondary liability.[241]

In the context of this Paper, likely claims for victims and the immediate families of victims of the race-based hate crimes mentioned above, align with those advanced in *Gonzalez*[242]. Although the plaintiffs' case in *Gonzalez*[243] was previously unsuccessful, the case is currently being reviewed by the Supreme Court.[244] Unlike the facts in *Gonzalez*[245], the facts of both the mass murders committed by Dylann Roof and Payton Gendron, were inextricable from pervasive racial stereotypes engrained in American history, that escalated to violent hatred.[246]

---

[232] *Id.*
[233] *Gonzalez v. Google, LLC*, 2 F.4th 871 (9th Cir., 2021).
[234] *Gonzalez v. Google, LLC*, 2 F.4th 871, 880.
[235] *Id.*
[236] *Id.* at 887.
[237] *Id.*
[238] *Id.*
[239] *Gonzalez v. Google, LLC*, 2 F.4th 871, 897 (9th Cir., 2021).
[240] *Id.* at 891.
[241] *Id.*
[242] *Id.* at 871.
[243] *Id.*
[244] Justia Supreme Court Center, Gonzalez v. Google, LLC, Oral Argument, OYEZ (Feb. 21, 2023), https://www.oyez.org/cases/2022/21-1333.
[245] *Gonzalez v. Google, LLC*, 2 F.4th 871, *supra* note 233.
[246] *Family Members of Buffalo Mass Shooting Want Focus on Preventing Racial Violence, Not Death Penalty, supra* note 77.

This Paper has indicated that at the core of programming and social media coding, programmers create and teach algorithms to use "class labels,"[247] which make active and obvious racial distinctions or choices.[248] These algorithmic choices are documented, and have been shown to result in racial profiling, disparate impacts in employment, and "[systemic disadvantages in] protected classes."[249] Where §230 allows for corporate immunity from the third-party impacts of "content-neutral algorithms,"[250] many social media algorithms are not racially content-neutral. Both *Algorithms of Oppression*[251] and this Paper have cited instances where algorithms flag, categorize, and push content involving black individuals, in ways that "Americans" and "ISIS members" are not as easily visually or racially categorized.[252] The lower court's reading of §230 required that internet service providers needed to employ algorithms with content- specified purposes and operate with a motive to support third-party content purposes, beyond baseline economic gain.[253] The corporate motive threshold is a significantly high threshold to meet when assessing liability for mass deaths. Section 230 has set an unfortunate precedent for corporate immunity.

CONCLUSION

In current times, there are very narrow possible claims available to individuals against social media corporations, who have suffered tortious or criminal harms, rooted in algorithmic recommendation actions.[254] Although individuals may seek limited, work-related damages based in tort, the contemporary legal immunities afforded to internet service providers under §230, for violent harms affecting large groups, are broad and forgiving.[255] Section 230 must be modified to provide large-scale recourse to protect classes who have experienced violent harms from algorithmic wrongs.

Both the plaintiffs in *Gonzalez*[256] and Judge Berzon of the Ninth Circuit Court have advocated for an overhaul of §230, to better protect the rights of targeted communities who have become victims of the polarizing ideologies supported and facilitated by internet service providers.[257] Judge Berzon, in her concurrence in *Gonzalez*[258], called "for a more limited reading of §230 immunity…urging the court to reconsider its precedent en banc to the extent that it holds that §230 immunity extends to the use of machine-learning algorithms to recommend content and connections to users."[259]

---

[247] Barocas, *supra* note 186.
[248] *Id.*
[249] *Id.*
[250] *Gonzalez v. Google, LLC*, 2 F.4th 871, 887, *supra* note 233.
[251] Noble, *supra* note 5, at 1-229.
[252] *Id.*
[253] *Gonzalez v. Google, LLC*, 2 F.4th 871, 891, *supra* note 233.
[254] *See generally*, Selena Scola, et al. v. Facebook, Inc., *supra* note 212.; *and Gonzalez v. Google, LLC*, 2 F.4th 871, *supra* note 233..
[255] *Gonzalez v. Google, LLC*, 2 F.4th 871, *supra* note 233.
[256] *Id.*
[257] *Id.* at 879.
[258] *Id.* at 871.
[259] *Id.*

Specific changes to §230 limiting internet service providers' immunity, have already been developed and proposed by the Department of Justice.[260] These potential limitations appropriately hold entities, like social media corporations, liable for their "Bad Samaritan"[261] acts under a new §230(d), including those that, "(1) purposefully promote, facilitate, or solicit third- party content [and behavior] that would violate federal criminal law; (2) have actual knowledge that specific content it is hosting violates federal law; and (3) fail to remove unlawful content after receiving notice by way of a final court judgement."[262] The additions to the current §230 immunity would allow victims to seek redress for specific harms experienced, due to "Bad Samaritan"[263] algorithm actions that have improperly recommended or failed to restrict harmful content.[264] Of the harms listed in the proposal, terrorism most aptly applies to both the harms claimed in *Gonzalez*[265] and in the domestic hate crime terrorist cases discussed.[266]

These immunity changes, and more, are necessary in the present day, where racialized violence is encouraged and permanently saved on internet sources. Where the youth are active social media users, formulating their world views on popularized information and disinformation[267], social media corporations must be made aware of the responsibility that is tied to the global impact their platforms create on normalcy. In Dr. Noble's words, "[w]e cannot ignore the long-term consequences…of the harmful effects of deep machine-learning algorithms, or artificial intelligence, on society."[268] These corporations must be incentivized to formulate algorithms that understand and promote the true realities of all communities and stop misinformation *before* it morphs into mass murder.

---

[260] *See* Off. of the Att'y. Gen., Dept. of Just.'s Rev. of Sec. 230 of The Communications Decency Act of 1996, THE U.S. DEP'T. OF JUST. ARCHIVES, https://www.justice.gov/archives/ag/department-justice-s-review-section-230-communications-decency-act-1996; Section by Section, https://www.justice.gov/file/1319326/download; *and* Ramseyer Draft Legislative Reforms to Section 230 of the Communications Decency Act, https://www.justice.gov/file/1319331/download.

[261] Michael R. Pence, *Office of the Attorney General*, *in* OFF. OF THE ATT'Y. GEN. 1 (2020).

[262] *Id.*

[263] *Id.*

[264] *Id.*

[265] *Gonzalez v. Google, LLC*, 2 F.4th 871, *supra* note 233.

[266] Pence, *supra* note 261, at 1.

[267] *See Family Members of Buffalo Mass Shooting Want Focus on Preventing Racial Violence, Not Death Penalty, supra* note 77.

[268] Noble, *supra* note 5, at 132.