Dangerous Speech and Social Media
Uncharted Strategies for Mitigating Harm Online

Client Organization
Center for Strategic & International Studies

Advisor
Sohini Chatterjee

Contributors
James Courtright
Hanna Homestead
Isabelle Lee
Daniyar Yerseitovich Nuraldinov
Dan Poniachik
Jasmine Ramsey

COLUMBIA | SIPA
School of International and Public Affairs

Contents

INTRODUCTION

Over the last decade, social media has fundamentally changed the way people communicate and access information around the world. While social media platforms provide new forums for constructive dialogue and personal expression, they are also used to proliferate dangerous online speech. Dangerous speech is defined as "any form of expression (e.g. speech, text, or images) that can increase the risk that its audience will condone or commit violence against members of another group." The characteristics of dangerous speech are that it promotes fear, and is commonly false, causes indirect harm, and is aimed at groups based on a shared identity.[1]

Overall, the response of governments and social media companies to the problem of online dangerous speech has been inconsistent and stymied by concerns over infringement of free speech rights, lack of will or capacity, and obfuscation over who is accountable for monitoring online content. Social media companies such as WhatsApp, Facebook, and Twitter, in particular, have been criticized for failing to address their role in exacerbating identity-based conflict by providing an online platform for dangerous speech.

This report investigates how dangerous speech shared on social media has intensified identity-based violence in Brazil, the Philippines, and Nigeria. Each case study investigates the national social media landscape, the weaponization of social media, and provides an analysis of how dangerous speech has contributed to specific cases of identity-based violence within that country. The report concludes with recommendations for social media companies on ways they can reduce the likelihood that dangerous speech is promoted or shared on their platforms globally.

---

[1]Susan Benesch et al., "Dangerous Speech: A Practical Guide," Dangerous Speech Project, January 9, 2020, https://dangerousspeech.org/guide/.

Methodology: The Dangerous Speech Framework[2]

To correctly categorize speech as dangerous, both its content and context must be considered. The case studies in this report use a five-part qualitative framework to analyze dangerous speech involved in an identity-based conflict in each country. The framework is used to examine the message itself, the audience, the historical and social context of the message, the speaker, and the medium with which a speaker delivers a message. The medium analyzed in this report is online social media platforms.

Key takeaways about dangerous speech:
- Dangerous messages often contain tactics that attempt to justify violence against groups of people with dehumanizing rhetoric, accusations in a mirror (when one side frames violence as a necessary means to protect against greater harm that is imagined or exaggerated), and accusation of threats to group integrity or purity or threats against women and girls.
- Dangerous speech is more likely to inspire violence if the audience is already susceptible to such messages, which can be observed by studying the message's historical context. Specifically, a history of violence between the groups, competition between groups for resources like land or water, and social norms, laws, and policies that put one group at risk are factors that increase the likelihood that an audience will be receptive to dangerous speech.
- The influence or authority of a speaker can make their speech more dangerous, both because of reputation or by an ability to spread the message to new or broader audiences.
- It is not necessary that all five elements be equally relevant for speech to be considered dangerous.

Speech alone does not cause violence, yet it can increase the risk that violence or physical harm against certain populations will be committed or condoned. The Dangerous Speech Framework provides a baseline with which to understand the mechanisms and motivations for dangerous speech, as well as possible opportunities for intervention.

Note: The report was conducted in full by SIPA, with CSIS as a client, and does not represent the research or opinions of CSIS.

---

[2] Benesch et al.

BRAZIL
CASE STUDY

Brazil is the largest country in Latin America. With a population of 212 million people, it ranks sixth in the list of countries in the world in terms of population and accounts for a third of Latin America's population.[3] According to the International Monetary Fund, in terms of gross domestic product (GDP), Brazil is the ninth-largest economy in the world with a nominal GDP of $1.87 trillion.[4] However, during recent years, Brazilian society has been facing many challenges: economic hardships, rising crime, polarization, increased immigration, violence, and inequality.

As a result, Brazil's GDP per capita decreased, reaching 8,752.400 USD in December 2019, compared with 9,039.840 USD in December 2018.[5] When comparing total numbers of murders, Brazil holds 7th place with 30.5 murders per 100,000 people.[6] Immigration has also been increasing social tensions.[7]

At the same time, over the past four years, there has been a steep rise in violent incidents against Brazil's LGBTQ+ community. Several news outlets cite President Jair Bolsonaro's rise in the political map in 2018 as deeply correlated with these incidents. Indeed, he himself has said very provocative statements, which his followers have been keen to retweet and exaggerate all over social media. Some media outlets now characterize Brazil as the LGBTQ+ murder capital of the world.[8]

This case study investigates the correlation between the rise of social media users and violence against the LGBTQ+ community in Brazil. Part one gives the latest data on the development of social media and the Internet in Brazil as well as legislation and state bodies responsible for the regulation in this field. Part two describes reasons and forms of using social media and impacts on drivers of dangerous speech in Brazil's internet. Part three provides a take on Bolsonaro's rise to power and his most divisive words against the LGBTQ+ community. Part four concludes this report.

---

[3] "Brazil Population 2020," WorldOMeter, 2020, https://www.worldometers.info/world-population/brazil-population/.

[4] Caleb Silver, "The Top 20 Economies in the World," Investopedia, March 18, 2020, https://www.investopedia.com/insights/worlds-top-economies/.

[5] "Brazil GDP per Capita [1962 - 2020] [Data & Charts]," CEIC, 2019, https://www.ceicdata.com/en/indicator/brazil/gdp-per-capita.

[6] "Global Study on Homicide" (Vienna, Austria: United Nations Office on Drugs and Crime, 2019).

[7] Shari Wejsa and Jeffrey Lesser, "Migration in Brazil: The Making of a Multicultural Society," Migration Policy Institute, March 29, 2018, https://www.migrationpolicy.org/article/migration-brazil-making-multicultural-society.

[8] Tyler Strobl, "Brazil as World LGBT Murder Capital and Rio's Place in the Data," Rio On Watch, July 10, 2017, https://www.rioonwatch.org/?p=37249.

PART ONE: Brazil's Social Media Landscape

1.1 Users

The development of the Internet in Brazil is encouraged by governments at the federal, regional, and municipal levels. Telecommunications companies in large and medium-sized cities, as well as small Internet providers, play an important role in increasing their accessibility to the population, including in remote areas of the country.

Since June 1990, when Brazil connected to the Internet, the number of Internet users has been growing rapidly. As of 2019, Internet penetration has reached 70 percent in Brazil, above the global average of 57 percent. More than 149 million out of the country's nearly 212 million inhabitants are internet users and 85 percent of internet users in Brazil browse the web every day.[9]
In addition, the growth in sales of mobile devices contributed to the active use of the Internet. Brazil has made a huge leap in the amount of data transmitted using wireless technology. Due to the significant rise in broadband connections, Brazilians have started using smartphones to access the internet. As of October 2016, 115 million users had 3G services.[10]

In general, Brazilians spend a total of nine hours 29 minutes per day on the Internet taking second place in the world (following The Philippines) and in terms of using the Internet from mobile devices Brazilians rank third globally (behind Thailand and The Philippines) spending four hours and 45 minutes daily.[11] Brazilians generally use mobile phones to access the Internet more than other devices. So, 66 percent of the population go online from smartphones and 63 percent of Brazilians use pre-paid mobile phones. This places the country under the global average of 75 percent.[12]

Most Facebook and Instagram users are under 54 years of age, with the highest concentration being of users below 35 years old. Male users tend to use it more than their female counterparts (Figure 1.1).
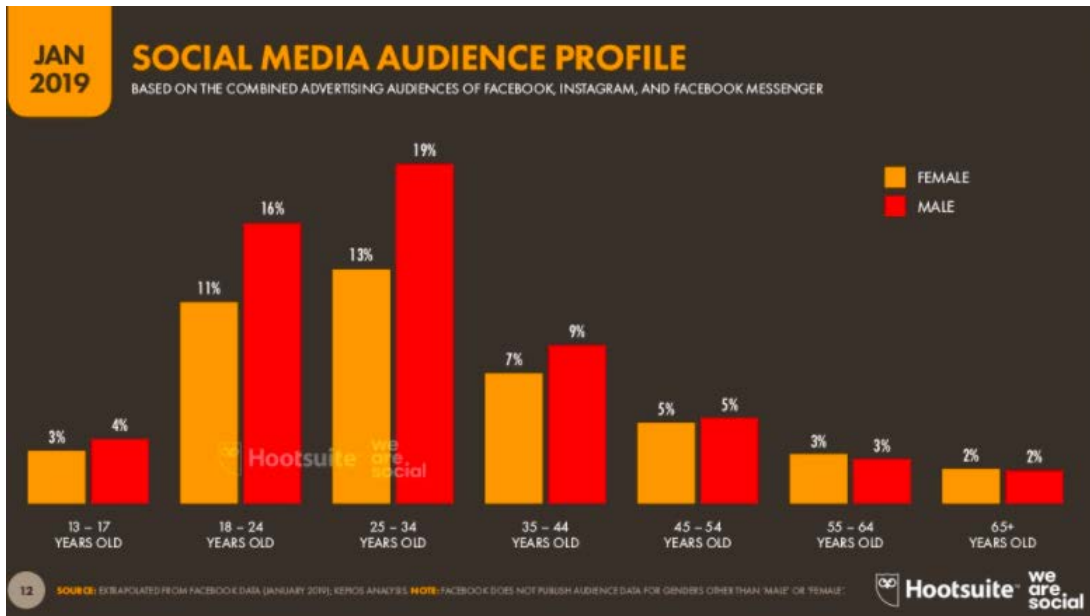
---

[9] Simon Kemp, "Digital 2019: Brazil" (Hootsuite and WeAreSocial, 2019), https://datareportal.com/reports/digital-2019-brazil.

[10] "Freedom on the Net 2019 - Brazil," Freedom House, 2019, https://freedomhouse.org/country/brazil/freedom-net/2019.

[11] Kemp, "Digital 2019."

[12] Kemp.

Figure 1.1[13]

1.2 Platforms

According to the Globalwebindex as of January 2020, the most-used social media platforms in Brazil are YouTube (96 percent), Facebook (90 percent), WhatsApp (88 percent), Instagram (79 percent), Facebook Messenger (66 percent), and Twitter (48 percent).

In a country famous for its telenovelas and television content, Brazilians actually spend more time using the Internet than watching television. In other words, the main sources of news for Brazilians is social media. As of 2019, in terms of total time spent on social media, 81 percent of Brazilians aged 13 and above are active on social media, against 58 percent worldwide. Considering the entire population, social media penetration is 66 percent.[14] According to this indicator Brazil takes second place in the world as Brazilians each day spend three hours and 34 minutes on social media.

---

[13] Kemp.

[14] Kemp.

Brazil is also the second Twitter user in the world.[15] Twitter also has a marketing, business development, and sales team located in Sao Paulo. Twitter has also recently enforced community guidelines removing posts of senior figures in Brazil.[16]

Figure 1.2
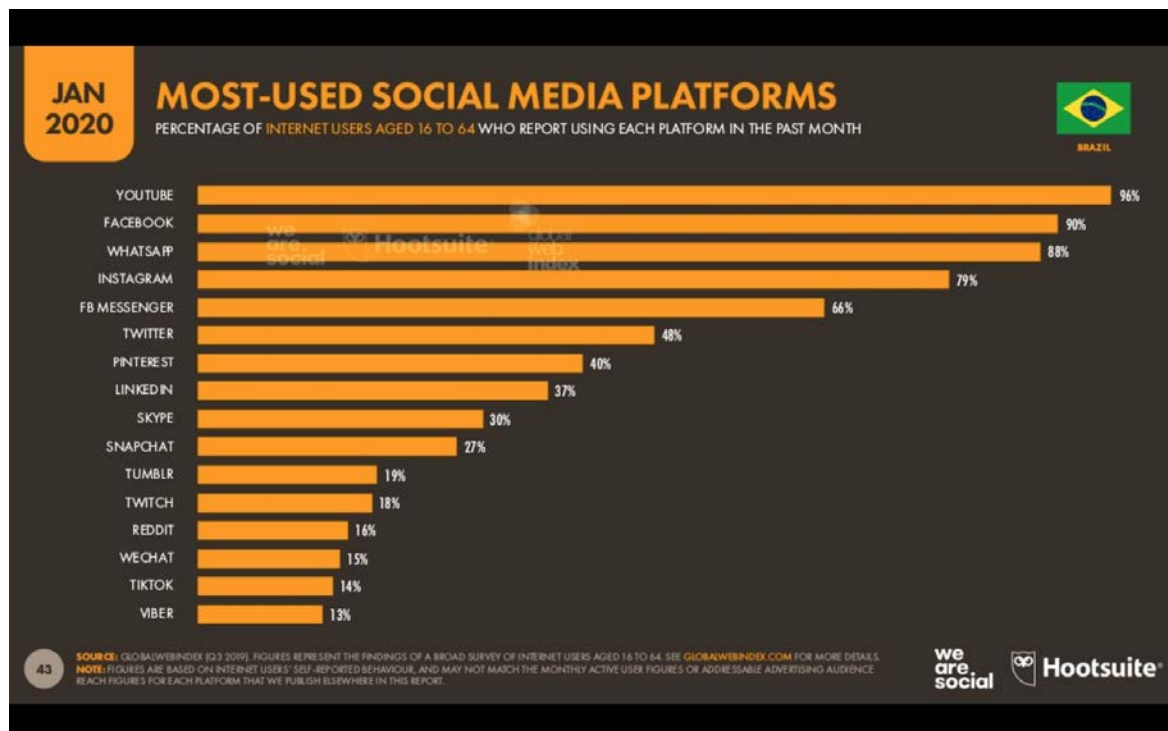
Source: Globalwebindex (2020)

WhatsApp is also very prevalent in Brazil. Fifty-three percent of WhatsApp users cite the application as the main source of news. Moreover, 58 percent of all WhatsApp users in Brazil are in groups with strangers to exchange information, and 18 percent of these users discuss news and events in these groups.[17] The facts stated above make Brazil one of the most social media heavy users in the world.

---

[15] "Brazil's Booming Social Media Revolution: Trends and Growth," Kwintessential, March 28, 2014, https://www.kwintessential.co.uk/blog/localisation/brazils-booming-social-media-revolution-trends-growth.

[16] Ryan Broderick, "Twitter Deleted Two Tweets From Brazil's President Jair Bolsonaro For Spreading Coronavirus Misinformation," BuzzFeed News, March 30, 2020, https://www.buzzfeednews.com/article/ryanhatesthis/twitter-deleted-tweets-brazil-president-coronavirus.

[17] Carolina de Assis, "Use of Instagram and WhatsApp for Online News Consumption Grows in Argentina, Brazil, Chile and Mexico: Reuters Institute," Knight Center for Journalism in the Americas,

1.3 Laws: Brazil Legal Framework

The rise of internet users in the country has led to new laws and relevant legal reforms during recent years.

In 2014, a law was passed in Brazil on the procedure for the legal regulation of the use of the Internet - Marco Civil da Internet. The procedure for using the Internet includes: the rights and guarantees of Internet users; provisions on access to the Internet and to Internet applications; rules of network neutrality; confidentiality of information transmitted over the Internet; regulates the procedure for the collection and storage of personal data; establishes the judicial procedure for the liability of Internet operators, and also establishes the forms and limits of liability of Internet operators.

Additionally, in 2018, the Brazilian Parliament approved the General Data Protection Law. This law guarantees the integrity and confidentiality of the exchange of information transmitted over the Internet; the inviolability and confidentiality of private messages stored by the user; as well as a guarantee of a ban on disclosing to third parties the personal data of users, including data on user connections and his access to Internet applications.

Moreover, in spite of provisions on the protection of the right to privacy, including the secrecy of correspondence, telegraphic, telephone, and data communications, in the Brazilian Federal Constitution of October 1988, Brazil has joined the International Covenant on Civil and Political Rights (ICCPR) and the American Convention on Human Rights or "Pact of San José de Costa Rica" (the "American Convention").[18] Both documents provide protection against unlawful interference and attacks on the right of privacy.

At the same time, in order to intercept communications Brazilian authorities can appeal to Law 9.296/96 which allows the interception of telephone communications and information technology systems, of any nature, for evidence in criminal investigation and in criminal procedural instructions.[19]

July 21, 2019, https://knightcenter.utexas.edu/blog/00-20991-use-instagram-and-whatsapp-online-news-consumption-grows-argentina-brazil-chile-and-me.

[18] "State of Privacy Brazil," Privacy International, January 26, 2019, http://privacyinternational.org/state-privacy/42/state-privacy-brazil.

[19] "Lei Nº 9.296, de 24 de Julho de 1996," Pub. L. No. LAW 9.296 / 1996 (1996), http://www.planalto.gov.br/ccivil_03/leis/L9296.htm.

In addition to the laws governing the interception of communications on the Internet, in 2012, the Brazilian parliament passed the Azeredo Law. According to its provisions, the country's authorities can create special teams to combat cybercrimes, as well as racist content on the Internet.[20]

In 2018, by the then president Michel Temer to oversee the implementation of these laws was created the so-called National Data Protection Authority (ANPD). In addition, there are two other agencies in Brazil responsible for regulating information and communication technologies in the country - the Brazilian Agency of Telecommunications (Anatel) and the Administrative Council for Economic Defense (CADE). The latter is also responsible for complying with competition rules in the telecommunications market.

PART TWO: Violence in Brazil has been increasing

Generally speaking, rates of violence in Brazil are historically peaking. On average, seven people died per hour in 2016—a staggering statistic that is partially driven by economic instability and dually influenced by political decisions to cut funding for national violence reduction programs. The prevalence of violence is positively correlated with regional socio-economic levels, particularly in northern, rural areas of the country.[21] In addition, areas with high rates of marginalization and large numbers of households consisting of single working mothers with young children experience higher rates of homicide.[22]

Homicide Rates for 2011 in Brazil's Municipalities

According to the report of the Brazilian Forum on Public Security (FBSP), the economic situation in the country is one of the main factors affecting the growth of homicides in Brazil. The widespread poverty in the country pushes people to crimes. Despite the increase in public funds allocated for security, the situation in Brazil is also complicated by the actions of the leadership. Therefore, the director of the FBSP Renato Sérgio de

---

[20] Luiz Piauhylino, "Azeredo Law PL 84/1999," Pub. L. No. PL 84/1999 Entire (1996), https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=15028.

[21] Catherine Osborn, "Why Brazil's Violence Keeps Getting Worse," NPR, November 2, 217AD, https://www.npr.org/2017/11/02/561505675/why-brazils-violence-keeps-getting-worse.

[22] Matthew C. Ingram and Marcelo Marchesini Da Costa, "Targeting Violence Reduction in Brazil: Policy Implications from a Spatial Analysis of Homicide" (Washington DC, USA: Latin America Initiative Foreign Policy at Brookings, October 2014), https://www.brookings.edu/wp-content/uploads/2016/06/Ingram-Policy-Brief.pdf.

Lima in 2018 during the election campaign compared Bolsonaro with the President of the Philippines, calling him the Brazilian version of Philippines' Rodrigo Duterte and stated that "Brazilian society is extremely violent," adding that with Bolsonaro urging "the chance of violence exploding is very high".[23]

Figure 2.1
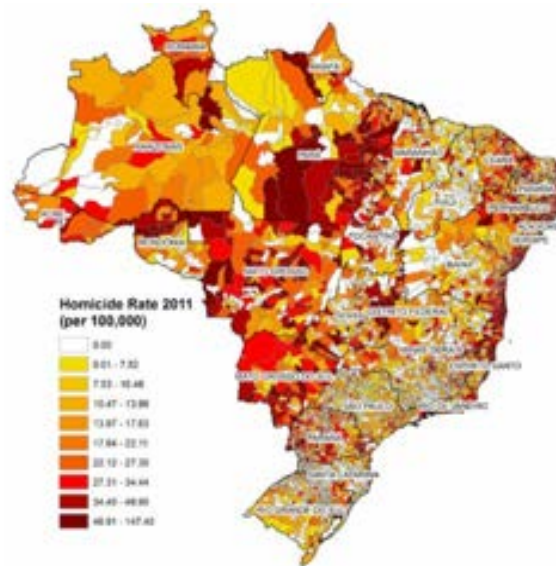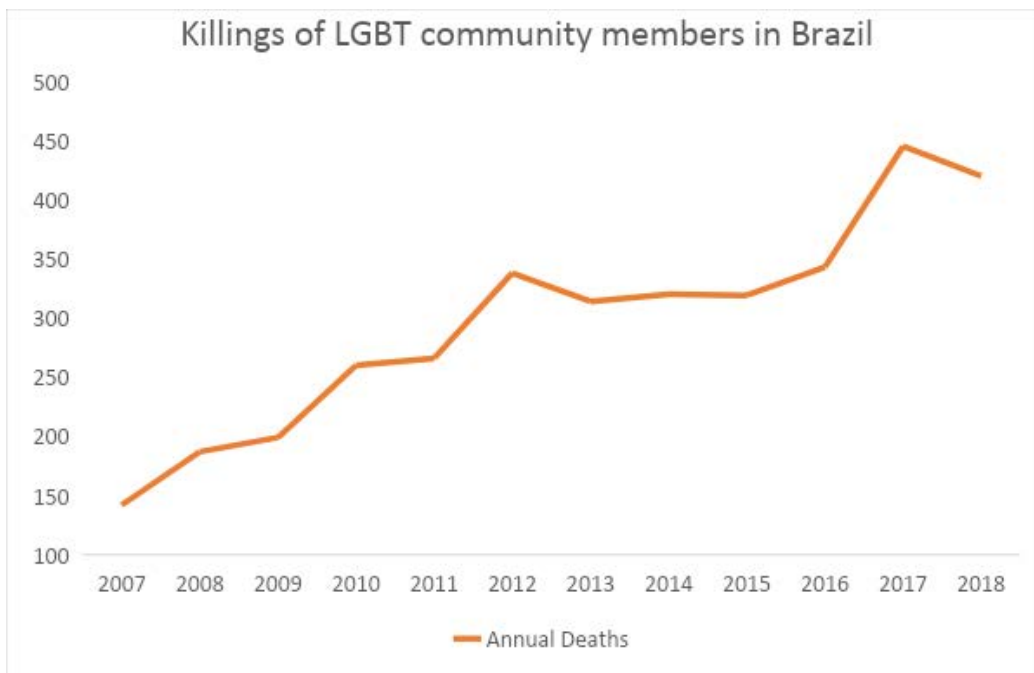

Source: The Brookings Institute (2016)

In particular, during recent years, Brazil has gained a reputation as the capital in the number of killings of LGBTQ+ community. According to the statistic in this field, the number of murders of LGBTQ+ people in Brazil is growing at an alarming rate. For example, the watchdog group Grupo Gay de Bahia in its release contended that at least 445 LGBT Brazilians became victims of homophobic crimes in 2017, comparing this figure to 2016, it increased by 30 percent[24] In addition, Grupo Gay de Bahia has identified that one gay individual is murdered per day in Brazil, with murders in Bahia accounting for 17 percent and in Sao Paulo for 10 percent of this total. This statistic has increased over time, and, according to different presidential periods; Lula saw 127 LGBTQ+ people killed each year, Dilma – 296, Temer – 407, and these numbers have only been increasing since Bolsonaro assumed power.[25]

---

[23] Andres Schipani and Joe Leahy, "Brazil Murder Rate Takes Its Toll on Economy," Financial Times, June 12, 2018, https://www.ft.com/content/d8e90bec-6dca-11e8-92d3-6c13e5c92914.

[24] Sam Cowie, "Violent Deaths of LGBT People in Brazil Hit All-Time High," The Guardian, January 22, 2018, https://www.theguardian.com/world/2018/jan/22/brazil-lgbt-violence-deaths-all-time-high-new-research.

[25] Cowie.

Figure 2.2

Killings of LGBT community members in Brazil

Source: Grupo Gay de Bahia (2019)

The statistics are staggering; every two hours, a Brazilian is beaten for their sexual orientation or gender identity, and every 19 hours a person is murdered.[26] During the 2018 election, reports of violence towards the LGBTQ+ community soared. Over the course of three days following the first round of elections on October 7th, 2018, 15 reports of physical, verbal, and cyber violence directed towards the LGBTI+ community were recorded in cities across Brazil.

Another prevalent and impactful example of violence against a member of the LGBTQ+ community during this critical time period before the 2018 elections concerned a 19-year-old woman from Porto Alegre, Brazil. In mid-October, pictures of a swastika brutally slashed onto a woman's stomach began to surface on the internet. Reports conclude that on October 8, the victim was attacked by three men who were allegedly prompted to incite violence against her in reaction to the shirt she was wearing at the time, which had "Elle Nao" written on the front, a reference to a Brazilian women's movement in opposition of President Jair Bolsonaro.

---

[26] Márcio Albino, "Eleições 2018 – 2° Turno – Posicionamento," Aliança Nacional LGBTI, October 10, 2018, https://aliancalgbti.org.br/2018/10/10/eleicoes-2018-2o-turno-posicionamento/.

The woman also adorned a sticker that said "#EleNão (Not Him)" with a rainbow, which is a well-known LGBTI+ symbol in Brazil. The victim reported to police that the men punched her in the abdomen and then two held her down while the third used a pocketknife to draw a swastika on her stomach. In late October, she was indicted for false crime reporting by the Rio Grande do Sul Civil Police. They concluded their investigation of the incident on October 24th, after which they decided upon reviewing security camera footage that the victim had misreported the incident. The chief delegate of the 1st Porto Alegre Police Station, Paulo Jardim, claimed that the symbol referred to Buddhism, signifying harmony, love, peace, and fraternity. The victim decided not to proceed with the case due to emotional shock. Gabriela Souza, the victim's lawyer, remarked that "society always relativizes the word of a woman who claims to be a victim."[27]

The situation is especially grim for members of the Trans community. According to the results of Trans Murder Monitoring (TMM) research project, between October 2018 and September 2019, the organization registered 331 cases of reported killings of trans and gender-diverse people around the world. The majority of the murders occurred in Brazil (40 percent), followed by Mexico (19 percent), and then the United States (9 percent).[28] This means that the likelihood of a transsexual being murdered is 10 times higher in Brazil as compared to in the United States.[29] This has greatly decreased the life expectancy of trans citizens of Brazil from 75 years for the average citizen to 35 years.[30]

Last year, the National Association of Travestis and Transsexuals (Antra) and Brazilian Institute Trans of Education (Ibte) published a dossier on murders and violence against trans people in Brazil. According to findings in the report, in absolute numbers, Rio de Janeiro was the largest killer of the trans population in 2018 with 16 murders. Rio de Janeiro is followed by Bahia with 15 cases, São Paulo with 14 cases, Ceará with 13 cases, Para with 10 cases, Minas Gerais with nine cases, Rio Grande do Sul, Paraná and Mato Grosso with eight cases, Pernambuco with seven cases, and Rio Grande do

---

[27] Felipe Souza, Ingrid Fagundez, and Matheus Magenta, "O que se sabe sobre o caso da suástica marcada em mulher de Porto Alegre," BBC News, October 25, 2018, https://www.bbc.com/portuguese/brasil-45767481.

[28] "TMM Update Trans Day of Remembrance 2019," Trans Murder Monitroing, November 11, 2019, https://transrespect.org/en/tmm-update-trans-day-of-remembrance-2019/.

[29] Henrique Mota, "Brazil: No Heaven for LGBT People," Medium, May 30, 2018, https://medium.com/@henriquemota/brazil-no-heaven-for-lgbt-people-df6d85c15fd1.
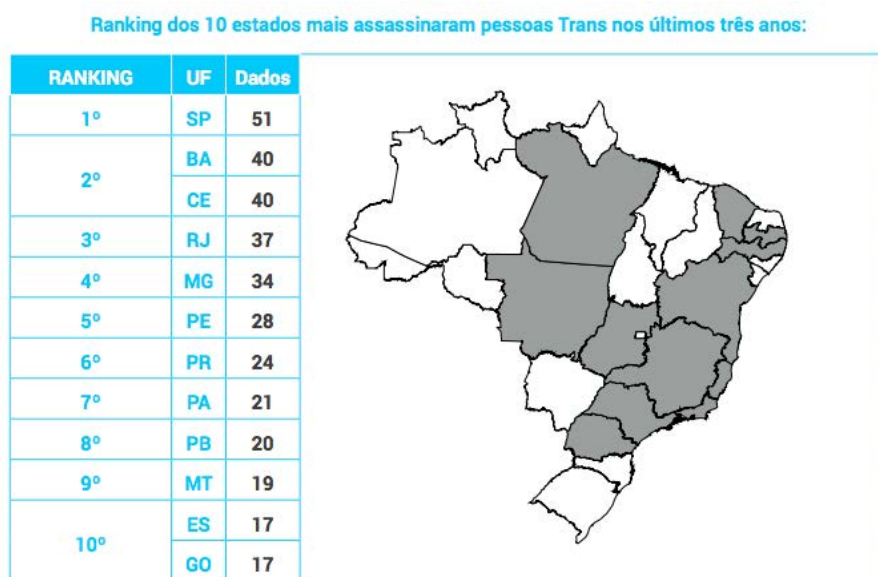
[30] Priscilla Moraes, "Brazil: Targeting Trans People With Impunity," Al Jazeera, April 22, 2015, https://www.aljazeera.com/indepth/features/2015/04/brazil-targeting-trans-people-impunity-150413210248222.html.

Norte, Goiás, Espírito Santo and Amazonas with 6 cases.[31]

If we combine the Brazilian states according to the regions of the country, the northeastern region (59) will be the first in terms of the number of murders, the southeast region (45) will follow it, then the southern region (20) and the northern region (19), and closes the list the central-west region (18). Unfortunately, of the total reported murder cases in 2019, only 11 had suspects identified by the police[32]

Ranking of the 10 States That Kill More Trans People in the Last Three Years

Figure 2.3

Ranking dos 10 estados mais assassinaram pessoas Trans nos últimos três anos:

RANKING	UF	Dados
1º	SP	51
2º	BA	40
2º	CE	40
3º	RJ	37
4º	MG	34
5º	PE	28
6º	PR	24
7º	PA	21
8º	PB	20
9º	MT	19
10º	ES	17
10º	GO	17

Source: ANTRA (2019)[33]

[31] Bruna G. Benevides and Sayonara N.B. Nogueira, "Murder and Violence against Travestis and Trans People in Brazil - 2018," trans. Sara Wagner York and Jacob R. Longaker (Brasília, Brazil: National Association of Travestis and Transsexuals, Brazil and Brazilian Institute Trans of Education, 2019), https://www.washingtonblade.com/content/files/2019/09/murders-and-violence-against-travestis-and-trans-people-in-brazil-2018.pdf.

[32] Andréa Martinelli, "Número de Pessoas Trans Assassinadas No Brasil Cai 24% Em 2019, Mas País Ainda é o Que Mais Mata," HuffPost Brasil, January 29, 2020, https://www.huffpostbrasil.com/entry/mortes-pessoas-trans-2019_br_5e309f6cc5b6e8375f6436a1.

[33] Bruna G. Benevides and Sayonara N.B. Nogueira, "Assassinatos e Violência Contra Travestis e Transexuais Brasileiras Em 2019" (Brasília, Brazil: National Association of Travestis and Transsexuals, Brazil and Brazilian Institute Trans of Education, 2020), https://antrabrasil.files.wordpress.com/2020/01/dossic3aa-dos-assassinatos-e-da-violc3aancia-contra-pessoas-trans-em-2019.pdf.

In sum, we can see that the levels of violence towards the LGBTQ+ community has been alarmingly rising during the last few years, and the trans community has been the hardest hit. The statistics of violence against the LGBTQ+ community are deeply disturbing; based on recorded death data, they are the highest in the world.

2.1 Social Media as a Weapon

Parallel to the widespread use of the Internet, representatives of various movements began to use social media to promote their ideas and expand their supporter groups. In this regard, social media such as Facebook and Twitter continue to play a key role in mobilizing the population and are actively used by civic activists.

One key example that shows this phenomenon occurred during the election campaign, in which social media was weaponized in ways that certainly affected the LGBTQ+ community. As mentioned in section 3.4, Bolsonaro's campaign accused opposer Haddad of providing gay kits in schools, and this false story was widely shared in WhatsApp groups.[34]

2.2 Drivers of Dangerous Online Speech

i. Deep Rooted Historical and Religious Reasons

Discrimination against homosexual behavior specifically can be dated back to the Spanish Inquisition, a historical source of Catholic doctrine in Brazil. Homosexuality was vastly believed to evoke plagues and famines during this time period.[35] Religious rhetoric in the country intends to discourage homosexuality and hateful behavior directed towards individuals concurrently by endorsing the sermon "do not hate the sinner but hate the sin," which may simultaneously contribute to the absence of constitutional laws criminalizing homosexuality and the prevalence of seemingly condoned bigotry, primarily stemming from radically conservative groups within Brazilian society and government.[36]

---

[34] Ciara Long, "False News Has Haunted Brazilian Elections for Years. This Time, It Was Worse.," Slate Magazine, October 30, 2018, https://slate.com/technology/2018/10/brazil-election-fake-news-whatsapp-facebook.html.

[35] Adrienne Rosenberg, "The Brazilian Paradox: The Lesbian, Gay, Bisexual, and Transgender Battle for Human Rights," in *Topical Research Digest - Revisiting Human Rights in Latin America*, ed. Arianna Nowakowski (Denver, USA: University of Denver, 2009), 16–30.

[36] Martin Nesvig, "The Complicated Terrain of Latin American Homosexuality," *Hispanic American Historical Review* 81, no. 3–4 (August 1, 2001): 689–729, https://doi.org/10.1215/00182168-81-3-4-689.

In relation to implicit chauvinistic cultural characteristics, various experts hypothesize that liberal government policies are not in congruence with social values that remain conservative and traditional. According to the statistics, the main religion in Brazil is Christianity, and around 90 percent of the population is either Catholic (65 percent) or Protestant (22 percent). In fact, Brazil has the world's largest Catholic population of 123 million people.[37]

In Brazil, Evangelical Christianity exported from the United States is gaining in popularity, a sect known for its hyper-orthodox conventions, including a strident opposition to homosexuality. Approximately a fourth of Brazil's population identified as Evangelical in 2016, a five percent increase from 1970, comprising a demographic shift that has extensively impacted the social and political fabric of Brazil. Not only have Evangelical churches influenced millions of Brazilians by purchasing radio and television time to further indoctrinate the population, these religious institutions have also propagated influence over the political sphere, inseminating conservative beliefs into political platforms and law-making processes.

Since 2010, more than 60 Evangelical lawmakers have been elected to the 513-seat lower Congressional house, marking a 50 percent increase in powerful conservative voices contributing to the Brazilian government. Evangelical politicians founded the B.B.B (Bullets, Beef, and Bible) caucus, a group in ardent opposition of laws that punish discriminatory acts, hate crimes, and general violence against the LGBTI+ community in Brazil.[38]

ii. Polarization, Social Tensions and Apathy in Law Enforcement

Certain academic researchers claim that two rival forces are working synchronously, feeding off of increases in its opposition's power.[39] Brazilians are holistically becoming more accepting and socially liberal, a process that has catalyzed a reactionary response from intolerant, far-right groups. Conservative politicians exploit vitriolic rhetoric to stymie progress made towards accruing LGBTI+ legal rights and protections on a national scale. For example, during a televised presidential debate in 2014, presidential candidate Levy Felix labeled members of the LGBTI+ community as unfit

---

[37] "Brazil Population 2020."

[38] Andrew Jacobs, "Brazil Is Confronting an Epidemic of Anti-Gay Violence," *The New York Times*, July 5, 2016, https://www.nytimes.com/2016/07/06/world/americas/brazil-anti-gay-violence.html.

[39] Javier Corrales, "Brazil's Recognition of Same-Sex Unions," Americas Quarterly, May 16, 2011, https://www.americasquarterly.org/article/brazils-recognition-of-same-sex-unions/.

parents. Well-known evangelical congressman Marco Feliciano predicated that AIDS is "a gay cancer" and claimed that the legalization of gay marriage "puts civilization and traditional families at risk of destruction."[40]

These sentiments are echoed by apathy in law enforcement, including lack of training programs focused on the prevention of biased hate crimes. Police tend to further humiliate victims of hate crimes by treating them as criminals or refusing to record an incident as a hate driven or biased crime, even when clear evidence of homophobia is presented in relation to the attack.

In addition, transgender victims experience the highest rates of mutilation related to crime. One notable account of a hate crime involving laceration and torture was caught on video in 2015. Samba dancer Piu da Silva pleaded for her life while being tortured until she was stabbed and shot six times. Her attackers were never found or prosecuted. Law enforcement not only disparages emphasis on the biased nature of crimes against members of the LGBTI+ community, but additionally reduces punishments for the perpetrators of those crimes.[41]

PART THREE: Bolsonaro's Use of Dangerous Speech

3.1 Social/Historical Context

Brazilian society comprises two juxtaposing characteristics: a population deeply ingrained in its historically Catholic roots (boasting the largest modern-day Catholic population in the world) and a sexually free, liberal public that widely endorses acceptance of the Brazilian LGBTI+ community. In fact, Brazil became the 11th country in the world to deem homosexuality politically and socially acceptable through a federal legalization process in 1823.[42]

Interestingly enough, despite Brazil's socio-cultural behavioral trademarks – commonly defined by cordiality, sympathy, and affection – Brazil is contemporaneously known for being one of the world's most flagrant transgressors of violence against its LGBTI+

---

[40] Jacobs, "Brazil Is Confronting an Epidemic of Anti-Gay Violence."

[41] Jacobs.

[42] "Constituição Do Estado Do Ceará -Atualizada Até a Emenda Constitucional N° 81, de 26 de Agosto de 2014" (2014), https://www.tjce.jus.br/wp-content/uploads/2015/07/Constituicao_Estadual.pdf.

citizens.[43] Between 1980 and 2006, 2,680 gays and lesbians were killed as a result of hate crimes, yet only 10 percent of their murderers were given prison sentences.[44] Overall, 63 percent of Brazilian LGBTI+ persons experienced some form of discrimination, distinguishing Brazil as more dangerous for LGBTI+ persons than Europe or the United States.[45]

Inherent patriarchal tendencies constitute one conspicuous facet of Brazilian society that contributes to widespread brutality against openly gay, lesbian, bisexual, and transgendered Brazilians. Patriarchy is often accompanied by male-dominated sexual norms. Gender roles are strongly prescribed to sexual orientation within this context; individuals are placed into a dichotomous framework, labeled as either 'male' or 'not-male.'[46] Classically depicted cisgender men are considered the 'givers,' and everyone else in society is categorized as the '*bicha,*' a derogatory term for 'female' in Portuguese with a strong connotation linked to 'receiving.' Consequently, gay men in particular face discrimination for falling on the incorrect side of the dichotomy in relation to their gender at birth.[47]

3.2 Bolsonaro's rise to power

Jair Messias Bolsonaro, (born March 21, 1955, Campinas, Brazil), is a right-wing nationalist, law-and-order advocate, and former army captain who has openly expressed admiration for the military government that ruled Brazil from 1964 to 1985 multiple times.

After leaving the army in 1988, Bolsonaro was elected to a seat on the Rio de Janeiro city council in 1989. Two years later he won a seat representing Rio de Janeiro in Brazil's federal Chamber of Deputies that he would hold for seven consecutive terms. From his first term on, Bolsonaro repeatedly praised the era of military rule and called for its return. He also began establishing a reputation for outspoken advocacy of

---

[43] Marco Akira Miura et al., "Between Simpatia and Malandragem: Brazilian Jeitinho as an Individual Difference Variable," *PLOS ONE* 14, no. 4 (April 15, 2019): e0214929, https://doi.org/10.1371/journal.pone.0214929; Rosenberg, "The Brazilian Paradox."

[44] "Brazilian President Calls National LGBT Conference," PinkNews, December 2, 2007, https://www.pinknews.co.uk/2007/12/02/brazilian-president-calls-national-lgbt-conference/.

[45] Tom Phillips, "Brazil in a Lather over Soap's Gay Kiss," *The Observer*, October 30, 2005, https://www.theguardian.com/world/2005/oct/30/brazil.gayrights.

[46] Don Kulick, "The Gender of Brazilian Transgendered Prostitutes," *American Anthropologist* 99, no. 3 (1997): 574–85, https://doi.org/10.1525/aa.1997.99.3.574.

[47] "Position Paper on the Language of 'Sexual Minorities' and the Politics of Identity" (Rio de Janeiro, Brazil: Sexuality Policy Watch, July 2008).

deeply conservative positions on social issues and was branded by critics as a misogynist, homophobe, and racist.

As mentioned earlier in the report, he has made several controversial remarks, not only towards the LGBTQ+ community but also towards women and people of color. For instance, when a female member of the Chamber of Deputies called him a rapist, Bolsonaro responded by saying, "I wouldn't rape you because you don't deserve it." Later, having stated that he was not a rapist, he added that if he were, he would not rape the congresswoman in question because she was "not his type." Commenting on the descendants of African slaves who organized the quilombo communities, Bolsonaro also said, "They do nothing! They are not even good for procreation."[48]

Bolsonaro's fortunes changed when Brazil's political establishment spiraled downward in the second decade of the 21st century and the economy went along. At the beginning of the second presidential term of Dilma Rousseff of the Workers' Party, the country sank deeper into a recession that had begun in 2014 and became mired in what some observers characterized as Brazil's worst economic crisis since the turn of the 20th century.

Bolsonaro, then, mounted a populist campaign that sought to take advantage of Brazilians' widespread disenchantment with the political establishment and rampant corruption. Using his outsider status to his advantage, Bolsonaro cast himself as an anti-establishment insurgent candidate with little concern for political correctness. Indeed, Bolsonaro was soon labeled the "Trump of the Tropics." Bolsonaro also won the support of the country's considerable Evangelical Christian population with his steadfast opposition to abortion, and his championing of law-and-order policies appealed to Brazilians concerned with rising crime and violence.

On September 6, 2018, while campaigning in Juiz de Fora, Bolsonaro was stabbed by a would-be assassin. His wounds required lifesaving surgery, forcing him to campaign from a hospital bed and then at home thereafter. However, Bolsonaro had already established a strong presence on social media, attracting more than 5.2 million Facebook followers at the time and as many as one million viewers to some of his video posts.

---

[48] Sam Meredith, "Who Is the 'Trump of the Tropics?': Brazil's Divisive New President, Jair Bolsonaro— in His Own Words," CNBC, October 29, 2018, https://www.cnbc.com/2018/10/29/brazil-election-jair-bolsonaros-most-controversial-quotes.html.

This prolonged stay in the hospital spared Bolsonaro from television debates where he has, in the past, struggled with difficult questioning, allowing him to control his messages via cellphone videos and social media posts to his millions of Facebook followers.

After being elected, his campaign sought to be a platform of conservative and divisive rhetoric. According to Brazil's government official website: "He has defended the reduction of the age of criminal responsibility, the possession of firearms for good citizens, the right to legitimate defense, the assurance of legal certainty during police action, and Christian and family values. He is the sponsor of the printed ballot bill (currently pending in Congress), which he believes will contribute to the holding of more reliable and auditable elections in the country and is a staunch congressional advocate of the fight against corruption".[49]

3.3 Bolsonaro and social media

Bolsonaro is an avid user of Social Media, which has played a particular role during his campaign and presidency. Notably, when he was in the hospital due to his attempted attack, he relied only on social media to direct his campaign, avoiding any TV appearances until well into the second-round runoff.

Bolsonaro has been doing weekly Facebook live videos since the past year and has deeply discredited mainstream media. Several of Bolsonaro's tweets fall in line with an onslaught of hostile tweets aimed at the mainstream media. He has accused media outlets such as Globo, Veja and Metrópoles of being "irresponsible," and "dishonest," labeling them "villains," "liars" and of publishers of "Fake News."[50]

His official Twitter account has 6.5 million followers, his official Facebook account has 10 million likes, and his Instagram page 16 million followers. As a comparison, several mainstream outlets which he alludes to have less followers. For example, one of brazil's main websites, Globo, only has 6.3 million Twitter followers.

---

[49] "Biography President Jair Bolsonaro," Federal Government of Brazil, 2020, http://www.brazil.gov.br/government/president.

[50] Sophie Foggin, "Bolsonaro's Weaponised Social Media Accounts," Latin America Reports, March 12, 2019, https://latinamericareports.com/bolsonaros-weaponized-social-media/1342/.

Bolsonaro's audience and supporters are greatly varied and diverse, though have several commonalities. In essence, his main base of support is composed of those fed up with rising violence and people tired of the PT party, some are members of the Brazilian Business community, Evangelicals, and the Farming Lobby.

Evangelical Christians - who make up 29 percent of all Brazilians - were one of the groups who supported Bolsonaro in greatest numbers. According to the last pre-election Datafolha poll, 61 percent of evangelicals voted for him. Bolsonaro describes himself as Roman Catholic, but his Christian rhetoric and his slogan, "Brazil above everything, God above all" won over many in the evangelical community.

Brazil is home to several fascist groups as well, including legitimate political parties. For instance, the Brazilian Integralist Front and PSL (Social Liberal Party), which is the right-wing party that selected Bolsonaro as its 2018 presidential candidate.[51] Members of radical groups or affiliates of fascist parties such as this one read or hear or watch homophobic statements made by Bolsonaro and used the supremacy of the source as legitimization to act. In a sense, the government is directly condoning hate crimes, and radical groups take advantage of this sentiment.

3.4 Bolsonaro incites hatred against the LGBTQ+ Community

Bolsonaro has repeatedly sought to downgrade the Brazilian LGBTQ+ community with hurtful comments. His comments aim to disenfranchise the community, allowing him the support of a conservative base that has spearheaded his rapid ascension to political power.

In 2011 he claimed he would "rather have [his] son die in a car accident than have him show up dating some guy" and that he "would be incapable of loving a gay son."[52] He made various other comments that alluded to violence against the LGBTI+ community, including that "Brazilians do not like homosexuals," "most homosexuals are murdered by their respective pimps at hours when good citizens are already asleep," "if I see two

---

[51] Tomas Castanheira, "Fascist Group Carries out Terrorist Attack in Brazil," World Socialist Web Site, January 8, 2020, https://www.wsws.org/en/articles/2020/01/08/braz-j08.html.

[52] Eliane Brum, "How a Homophobic, Misogynist, Racist 'Thing' Could Be Brazil's next President," The Guardian, October 6, 2018, https://www.theguardian.com/commentisfree/2018/oct/06/homophobic-mismogynist-racist-brazil-jair-bolsonaro.

men kissing in the street, I will beat them," and "if children start to become a little gay, whip them to change their behavior.[53]

On October 4, 2018, Bolsonaro reverted again to his pre-presidential campaign stigmatization of the LGBTI+ community. He stated that "a father does not want to come home and see his son playing with a doll by the influence of school."[54] He was referring to school supplies as 'gay kits' with the potential to encourage children to become homosexual.

Bolsonaro changed his public stance on homosexuality in congruence with the beginning of his presidential campaign in mid-October in 2018. He reassured the Brazilian public that "we are going to make a government for everybody. For gays… it is a work for everyone."[55]

His previous reputation as a homophobic member of Brazil's government combined with slurs made prior to the election signaled to radical people and groups within Brazil's population that violence against LGBTI+ Brazilians would be accepted and validated by the administration. Despite Bolsonaro's slight change in rhetoric later in October 2018, the memories of his previous discriminatory statements stuck with the most radical factions included in Brazilian society.

Bolsonaro has made clear that he is not fond of the LGBTQ+ community. His statements before becoming President of Brazil were shocking, which he has not retreated since being elected. A few examples of his statements before being elected president are the following:[56]

---

[53] "Palmada Muda Filho 'Gayzinho', Declara Deputado Federal," Folha de S.Paulo, November 26, 2010, https://www1.folha.uol.com.br/fsp/cotidian/ff2611201025.htm.

[54] iG São Paulo, "Bolsonaro Sobre Homossexuais: 'Serão Felizes Se Eu for Eleito,'" Último Segundo, October 4, 2018, https://ultimosegundo.ig.com.br/politica/2018-10-04/bolsonaro-sobre-homossexuais-serao-felizes.html.

[55] Gustavo Maia and Luís Adorno, "Em última fala antes de votação, Bolsonaro faz aceno para ateus e gays," UOL, June 10, 2018, https://noticias.uol.com.br/politica/eleicoes/2018/noticias/2018/10/06/em-ultima-fala-antes-de-votacao-bolsonaro-faz-aceno-para-ateus-e-gays.htm.

[56] "Who Is Jair Bolsonaro? Brazil's Far-Right President in His Own Words," The Guardian, October 29, 2018, https://www.theguardian.com/world/2018/sep/06/jair-bolsonaro-brazil-tropical-trump-who-hankers-for-days-of-dictatorship.

"I would be incapable of loving a homosexual son. I'm not going to be a hypocrite: I'd rather my son died in an accident than showed up with some bloke with a moustache." (Playboy Magazine, June 2011)

"I won't fight it or discriminate, but if I see two men kissing each other in the street, I'll whack them." (October 2002)

"We Brazilians don't like homosexuals." (Interview with Stephen Fry, 2013)

"You have a terribly homosexual face" (to a Globo journalist, December 2019)

"Brazil Must not Become a Gay Tourism Paradise. "Brazil can't be a country of the gay world, of gay tourism. We have families"[57] (Crusoé Magazine, April 2019)

"The World Health Organization incentives child masturbation and Homosexuality" (Bolsonaro's Personal Facebook Account, April 2020[58])

His interview with British actor Stephen Fry in 2013 was particularly chilling. In this interview, he allowed himself to declare on behalf of all Brazilians that Brazilians do not like homosexuals. As a congressman, Bolsonaro openly opposed the law criminalizing homophobia and introducing anti-homophobia lessons in schools.

During the latest round of elections, several messages in WhatsApp depicted rumors about his main contestant, Haddad, about his campaign wanting to introduce gay kits for teaching children in schools. Through spreading fake news about his opponent, Bolsonaro pursued two goals, he wanted to denigrate Haddad and incite hatred of gays in Brazilian society.[59]

[57] Tom Phillips and Anna Jean Kaiser, "Brazil Must Not Become a 'Gay Tourism Paradise', Says Bolsonaro," The Guardian, April 25, 2019, https://www.theguardian.com/world/2019/apr/26/bolsonaro-accused-of-inciting-hatred-with-gay-paradise-comment.

[58] "Bolsonaro diz que OMS incentiva a sexualidade infantil, mas apaga post," BR Político, April 30, 2020, https://brpolitico.com.br/noticias/bolsonaro-diz-que-oms-incentiva-a-sexualidade-infantil-mas-apaga-post/.

[59] "Jair Bolsonaro Accused of Creating 'criminal Network' to Spread Fake News in Brazil Election," The Telegraph, October 18, 2018, https://www.telegraph.co.uk/news/2018/10/18/jair-bolsonaro-accused-creating-criminal-network-spread-fake/.

In March 2019, he condemned Brazil's Carnival through a post that had a link to a gay pornographic video. According to Intercept journalist Glenn Greenwald, who now lives in Brazil, the video Bolsonaro posted was "gay kink porn"[60] and was deeply disturbing as it was intended to incite homophobia.

In April 2019, Bolsonaro stated in an interview that Brazil must not become a gay tourist paradise[61]and in May that year, Bolsonaro accused the country's Supreme Court of "legislating" from the bench after a majority of its justices voted to make homophobia a crime like racism. He also cancelled in August funding for TV shows that had LGBT content in August, sharing in his official Facebook page the following:

Screenshot of Bolsonaro's Official Facebook Page

---
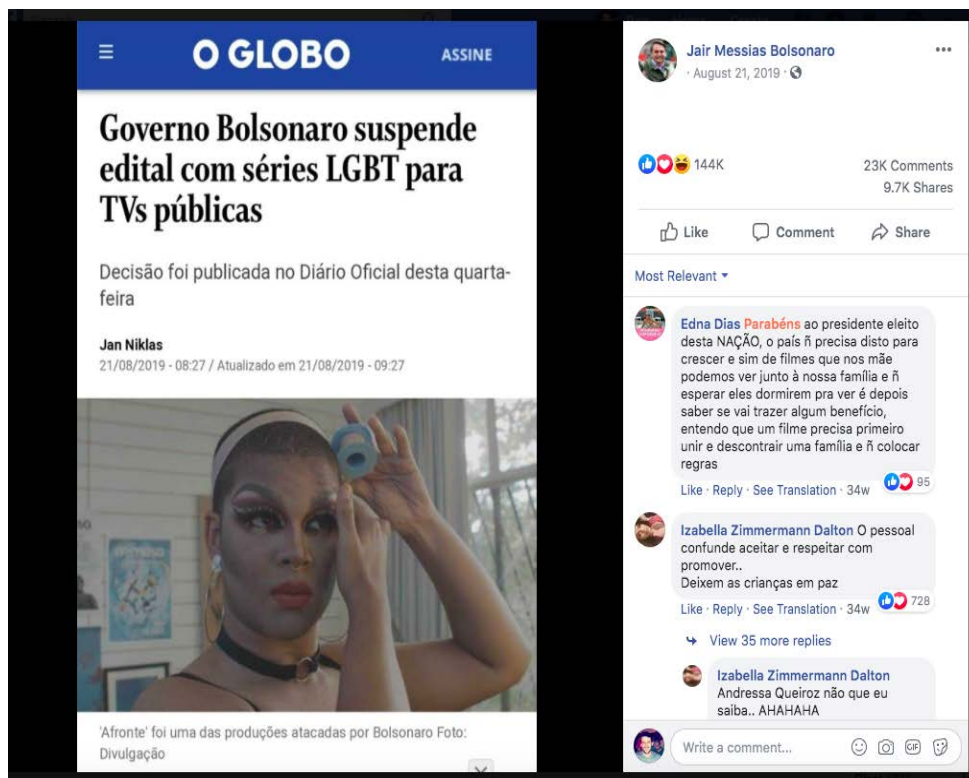
[60] Jen Kirby, "Brazil's Far-Right President Tweeted out a Pornographic Video to Condemn Carnival," Vox, March 6, 2019, https://www.vox.com/world/2019/3/6/18253070/jair-bolsonaro-tweet-golden-showers-pompeo.

[61] Phillips and Kaiser, "Brazil Must Not Become a 'Gay Tourism Paradise', Says Bolsonaro."

In a recent exchange of words with the World Health Organization, Bolsonaro on April 29 said in his official Facebook account that the Institution was "encouraging the masturbation and homosexuality of children." Bolsonaro then went back and deleted the post published on his Facebook profile minutes later, but several media accounts were able to screenshot the comments.[62] He was mocking the WHO guidelines for COVID, asking if we should "also follow their educational guidelines…"
The bulk of Bolsonaro's statements inciting violence against members of the LGBTQ+ community were made publicly by 2011. His rise to the top of the political hierarchy in Brazil reinforced these attitudes seven years later, even in light of last-minute statements he made to bolster popular support immediately before the commencement of the first round of elections in 2018.

During that electoral year, violence spiked, as we could have expected. As mentioned earlier, the woman from Porto Alegre who was attacked and got a swastika carved, was triggered by support of the feminist anti-Bolsonaro campaign and LGBTQ+ sticker she as wearing.[63]

As several news accounts from 2018 recall, several Transsexuals were murdered under calls of "Bolsonaro President"[64] as violence spiked that electoral month. The LGBTQ + National Alliance mentioned the 10th of October that killings were on the rise because of the electoral campaign, opened a national hotline to denounce this and said in a national declaration: "considering the notorious LGBTQphobic statements by candidate Bolsonaro, the feeling of "legitimation" that his candidacy has given to people and groups that have assaulted and killed minorities in recent times and his clearly authoritarian character, the National Alliance LGBTQ+ + expressly supports the Haddad / Manuela ticket in this second round"[65]

On December 12, 2019, Veronica Oliveira's death was one of the most notorious murders of the year, just days after she was crowned godmother of the LGBTQ Alternative Parade. Oliveria was a transsexual woman and businesswoman in a housing

---

[62] "Bolsonaro diz que OMS incentiva a sexualidade infantil, mas apaga post."

[63] Souza, Fagundez, and Magenta, "O que se sabe sobre o caso da suástica marcada em mulher de Porto Alegre."

[64] Lu Sudré, "Transexuais são assassinadas sob gritos de "Bolsonaro presidente"," Brasil de Fato, October 24, 2018, https://www.brasildefato.com.br/2018/10/24/transexuais-sao-assassinadas-sob-gritos-de-bolsonaro-presidente.

[65] Albino, "Eleições 2018 – 2º Turno – Posicionamento."

that has been housing trans women for ten years.[66] She is a leader of the community and has been a leader over the past 10 years.[67]

As reported by several NGOs and the LGBTQ+ community, cited in the report above, his statements and the timing of his election appear to be notably correlated with increased violence towards members of this community.

PART FOUR: Conclusion

Over the past decade, Brazil has seen an increase in crimes against members of the LGBTQ+ community, which has been clearly demonstrated by both statistical data and the influence of the statements of Bolsonaro before his election as president and after taking office. Being an ardent womanhater, homophobic, and a supporter of dictatorship, he repeatedly expressed his dislike for the LGBTQ+ community. With the help of modern technology and social media, he made statements that made it clear to the people of the country that non-traditional orientation in Brazil was unacceptable.

Given the growing number of Internet users among the country's population, Bolsonaro's messages quickly spread and found a response among his supporters. This was also facilitated by the fact that most Brazilians began to read the news more via the Internet than through traditional television and radio. It must also not be forgotten that a decrease in economic growth and an increase in unemployment and poverty also contributed to the growth of aggressive moods in society.

In addition, despite the existence of laws regulating the use of the Internet, they do not have provisions prohibiting dangerous speech online. The absence of such provisions did not allow regulating online publications. There is also no information that social media somehow regulated this issue.

In a situation where there is no proper control from the state when no one forces social media to regulate and monitor dangerous content on the network, Bolsonaro could and continues to make publications at his own discretion. Considering the fact that

---

[66] "Verônica, Ativista Trans Em Santa Maria, é Brutalmente Assassinada Com Uma Facada," Esquerda Diario, December 12, 2019, http://www.esquerdadiario.com.br/Veronica-ativista-trans-em-Santa-Maria-e-brutalmente-assassinada-com-uma-facada.

[67] Naion Curcino, "Líder do movimento LGBT+ de Santa Maria morre após ser esfaqueada," GaúchaZH, December 12, 2019, https://gauchazh.clicrbs.com.br/seguranca/noticia/2019/12/lider-do-movimento-lgbt-de-santa-maria-morre-apos-ser-esfaqueada-ck42qko1x04wf01rz1ac7d2ne.html.

Bolsonaro has a relatively large audience of followers, his messages spread very quickly. In addition, these messages could be transmitted further from group to group.

Also, it is important to note that Brazil is a Christian country with a large share of religiously minded people for whom non-traditional values are not always acceptable. In this situation, Bolsonaro's statements and posts contributed to the growth of homophobic sentiments in Brazilian society, as a result of which there was an increase in crimes and killings of LGBTQ+ people.

In this regard, we believe that in order to rectify the situation in Brazil, it is necessary to adopt amendments to the legislation governing the activities of social media and toughening the responsibility for dangerous speech online. This work should be carried out with the participation of representatives of the LGBTQ+ community, human rights activists, clergy, and journalists.

THE PHILIPPINES
CASE STUDY

The Philippines, an archipelagic nation situated in the Western Pacific Ocean, consists of over 7,500 islands with over 180 different languages. The country is diverse given its cultural and geographic divide, yet relatively religiously homogeneous, with 95 percent of Filipinos being Roman Catholic.

Faced with over four centuries of oppressive rule from various foreign nations, the state has a complicated social and political landscape that has experienced decades of smaller scale communal conflict since its independence from the Americans in 1946.

Today, the Philippines has been crowned the "social media capital of the world" by the annual Global Digital Report,[68] from 2016 to 2019 despite having among the poorest internet infrastructures in the world. Brazil came in second, followed by Thailand in 2019. An average Filipino spends at least four hours on social media alone and a total of nine hours surfing the web every day.[69]

Unlike many decolonized states, Philippine conflicts have largely been motivated by evolving ideological differences that are not only rooted in ethnic, religious, or even class divisions but due to political transitions, particularly election seasons. And as one of the oldest Asian democratic regimes, its weak multi-party electoral system has largely mitigated the potential for deeply polarized political ideologies.

To understand the social and historical contexts of dangerous speech and conflict as well as how these exacerbate pre-existing tensions, it is necessary to frame them in contemporary terms.

In a country where over 70 percent[70] of the 108 million population is connected to the internet, with nearly 100 percent having a Facebook account, the social media platform has connected millions of Filipinos across the world. Simultaneous to the efficient growth of social media, an unprecedented wave of conflict resulting in nearly 30,000 homicides occurred between 2016 to 2020, with spikes in violence coinciding with the spread of dangerous speech on the platform. This highlights a need to better

---

[68] Simon Kemp, "Digital 2019: The Philippines" (Hootsuite and WeAreSocial, 2019), https://datareportal.com/reports/digital-2019-philippines.

[69] Janvic Mateo, "Philippines Still World's Social Media Capital – Study," February 3, 2018, https://www.philstar.com/headlines/2018/02/03/1784052/philippines-still-worlds-social-media-capislow%20tal-study.

[70] J. Müller, "Philippines: Internet Penetration Rate 2017 to 2023," Statista, March 4, 2019, https://www.statista.com/statistics/975072/internet-penetration-rate-in-the-philippines/.

understand how social media is interacting with the pre-existing social and historical tensions in the state.[71]

This case study investigates the rapid rise of Facebook in the Southeast Asian nation and how it may have exacerbated violence in the country. Part one gives a brief overview of the country's political and social history and lays out the social media landscape in the Philippines. Part two details Facebook's strategy in permeating the nation as well as the existing laws and policies that concern social media regulation. Part three discusses the weaponization of Facebook in driving dangerous speech by the current administration, the direct effects of this on drug-related deaths, and the response of key stakeholders. Part four concludes this report.

PART ONE: Philippine's Social Media Landscape

Prior to internet access, the Philippines' media consumption was primarily through radio and television. Following decolonization from the Americans in 1946, media consumption and trends demonstrated a cultural tendency to embrace both technology and media.[72]

As of April 2020, the Philippines leads the world in time spent online and on social media. On average, Filipinos spend four hour and 12 minutes a day on social media compared with the global average of two hours and 16 minutes.[73] According to available data, three hours and 22 minutes of daily time on social media is spent in the Philippines watching videos on Facebook or YouTube. And since 2013, when the internet became more accessible to a wider set of the population, time spent on social media is growing at an annual rate of 16 percent in the state, compared to the global average of 13 percent.[74]

---

[71] Mari Chrys Pablo, "Internet Inaccessibility Plagues 'Social Media Capital of the World,'" The Asia Foundation, October 24, 2018, https://asiafoundation.org/2018/10/24/internet-inaccessibility-plagues-social-media-capital-of-the-world/.

[72] Pauline Estella and Matin Löffelholz, "Philippines - Media Landscapes" (Maastricht, Netherlands: European Journalism Centre, 2019), https://medialandscapes.org/country/philippines.

[73] Doris Dumlao-Abadilla, "At 10 Hrs, 2 Mins a Day, Filipinos Spend Most Time Online," February 1, 2019, https://business.inquirer.net/264438/at-10-hrs-2-mins-a-day-filipinos-spend-most-time-online.

[74] {Citation}

Most Filipinos access the internet and social media through smartphones,[75] with 72 million mobile social media users in the state, where smartphones outnumber the population at 116 percent or 124.2 million cellular subscriptions.[76]Among the Filipinos connected to the internet, 76 million are active social media users—a 71 percent social media penetration rate.

An estimated 10.2 million Filipinos live or work outside of the country and as a culture, family is a fundamental value to many. Geographic dislocation among family and friends, as such, contributes to widespread and increasing Facebook penetration in the culture.[77]

1.1 Users

Since 2016, Facebook penetration in the Philippines is growing on average by 2.6 percent each quarter with social media use on a steady rise with annual usership increasing. However, visits to Facebook per month average 423.3 million compared with the 247.3 million visits YouTube receives.[78]

On Facebook, the average user is 25 (the same as the median age in the Philippines), with 53 percent of users reported as female and 47 percent reported as male. Ninety-eight percent of users are reported to be reached by paid advertisements, of which have a quarterly growth rate of 2.7 percent as of January 2019.[79]

---

[75] Dan, "Social Media Statistics in the Philippines," Talkwalker, August 7, 2019, https://www.talkwalker.com/blog/social-media-statistics-philippines.
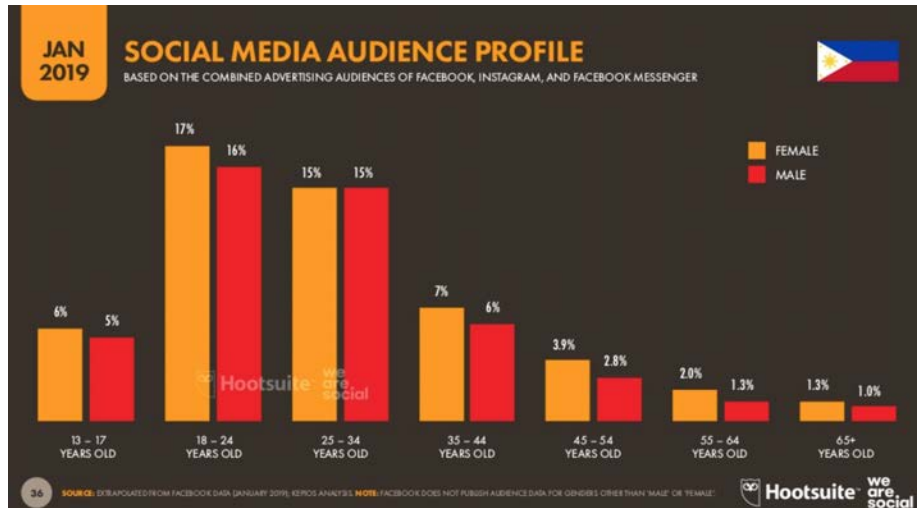
[76] Kemp, "Digital 2019: The Philippines," 2019.

[77] "Total Number of OFWs Estimated at 2.3 Million (Results from the 2018 Survey on Overseas Filipinos)," Republic of the Philippines Philippine Statistics Authority, April 30, 2019, https://psa.gov.ph/statistics/survey/labor-and-employment/survey-overseas-filipinos.

[78] Kemp, "Digital 2019: The Philippines," 2019.

[79] Kemp.

Figure 1.1

1.2 Platforms

Of those online, nearly 100 percent use Facebook—a striking figure for a country wherein the number of smartphones outnumber people.[80]

Figure 1.2

---

[80] Jake Swearingen, "Facebook Used the Philippines to Test Free Internet. Then a Dictator Was Elected," New York Magazine, September 4, 2018, https://nymag.com/intelligencer/2018/09/how-facebooks-free-internet-helped-elect-a-dictator.html; Lauren Etter, "What Happens When the Government Uses Facebook as a Weapon?," Bloomberg Businessweek, December 7, 2017, https://www.bloomberg.com/news/features/2017-12-07/how-rodrigo-duterte-turned-facebook-into-a-weapon-with-a-little-help-from-facebook.

Globally, the annual digital growth rate is by far the largest in the Asia-Pacific Region, where in 2019 social media use increased by 218 million users.[81] Compared with Europe, the Americas, the Middle East, and Africa, the region's digital growth is rapid. In 2019 alone, the region had more new social media users than all other regions by more than 175 million people.[82] The Philippines. in fact, leads the region in digital engagement, most specifically through social media usage. And, the average social media user has 11 accounts in the state. As evidenced in Figure 1.3:

Leading International Facebook Usage as of January 2020

Figure 1.3
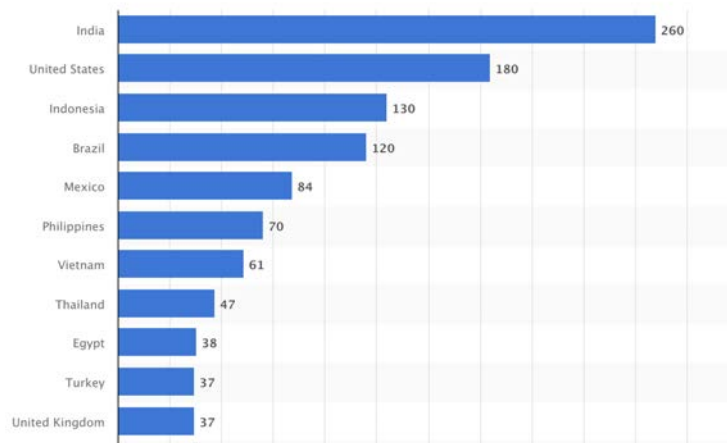
For the Philippines, Facebook is the internet and, by extension, the news. Still, other social media platforms' penetration rates are increasing, in particular, available data flags YouTube and LinkedIn as the second and third most quickly growing platforms in the country respectively.[83] As of January 2019, YouTube penetration is the second highest in the country (61 percent) and far above WhatsApp (43 percent), Facebook Messenger (41 percent), Instagram (38 percent), and Twitter (5 percent). Facebook, by far, remains the most popular and influential platform.[84]

---

[81] Simon Kemp, "Digital 2019: The World" (Hootsuite and WeAreSocial, 2019), https://datareportal.com/reports/digital-2019-philippines.

[82] The aggregate data on global digital growth rates is different between states. Although there is globally an average of 13 percent social media growth, this number is per capita. Therefore, in the Asia-Pacific region, the qualitative indicators can be misleading as the number is much higher in the area than others even though it is represented by the same statistic.

[83] "Social Media Statistics in the Philippines."

[84] Kemp, "Digital 2019: The Philippines," 2019.

Figure 1.4
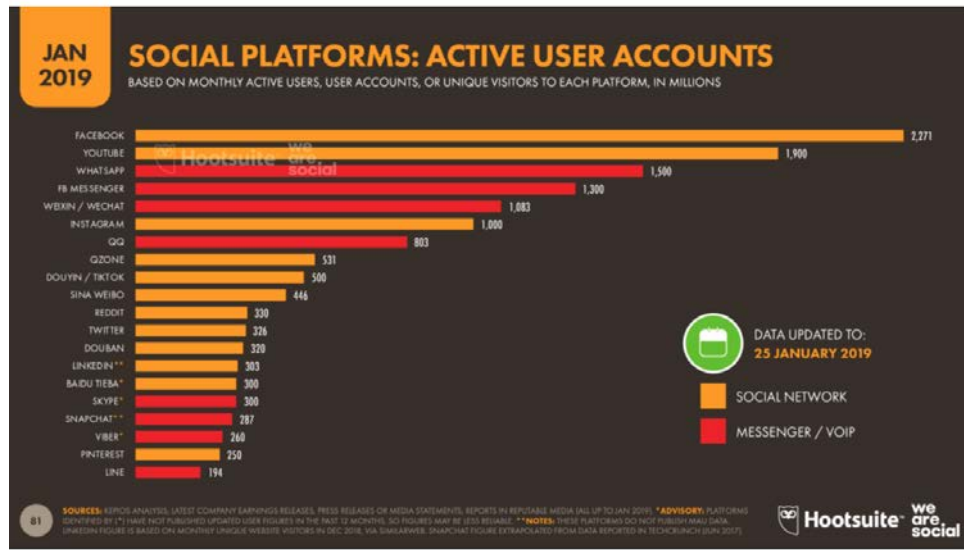
1.3 Laws: Philippine Legal Framework

Similar to countries around the world, the unprecedented impact of internet access has demanded a legal infrastructure related to the medium in the Philippines.

The Philippine domestic and international legal frameworks offer a multitude of protections and precedent per social media use at the intersection of speech, some of which were pre-existing. And, since the penetration of internet usership has grown, older legal systems have been reinterpreted to reflect digital rights and online speech.

For example, the state is party to all seven of the core international human rights treaties and most of their optional protocols, of which further bind the state to enforcement mechanisms. Namely, the Philippines is party to the International Convention on Civil and Political Rights (ICCPR) and both of its optional protocols.[85] With emphasis on negative rights, the state is obligated to refrain from illegitimate restrictions on freedom of expression: directly or indirectly. The ICCPR article 20.2 refers to the illegal nature of dangerous speech.[86]

---

[85] "Status of Ratification - Interactive Dashboard," United Nations Office of the High Commission on Human Rights, May 11, 2020, https://indicators.ohchr.org/.

[86] "International Covenant on Civil and Political Rights (ICCPR)" (1966), 20.2. Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law."

Further, the Human Rights Committee's General Comment 34, referring specifically to ICCPR Article 19 on freedoms of opinion and expression, emphasizes the need for party states to ensure diverse and independent media forms are both accessible, widely available, and that personal expression of opinion is decriminalized. Violations of articles 19 and 20 have been raised by human rights advocates internationally and in the country since 2016, of which these concerns were reflected in the state's most recent Universal Periodic Review.[87]

Domestically, the 1987 Constitution of the Philippines, Article III Section IV[88], has provided a basic framework for understanding the states obligations pertaining to freedom of expression in the context of dangerous speech online. And in important jurisprudence, social media use has been interpreted to garner some parameters from the Bill of Rights as it names the right against unreasonable searches and seizures; and, the right to privacy of communication and correspondence. This has been interpreted to social media use as protections for Filipinos to communicate freely and without interference, including surveillance or censorship.[89]

Still, there is no direct policy on social media privacy in the Philippines. There are, however, several laws including jurisprudence[90] concerning privacy as it relates to social

---

[87] "ICCPR General Comment 34" (2011).

[88] Excerpt: "Section 4. No law shall be passed abridging the freedom of speech, of expression, or of the press, or the right of the people peaceably to assemble and petition the government for redress of grievances."

[89] "The 1987 Constitution of the Republic of the Philippines - Article III," Official Gazette of the Republic of the Philippines, 1987, https://www.officialgazette.gov.ph/constitutions/the-1987-constitution-of-the-republic-of-the-philippines/the-1987-constitution-of-the-republic-of-the-philippines-article-iii/.

[90] Two primary cases of relevance: St. Theresa's College Cebu Case, referring to Facebook privacy and ruling stating that once images or videos are shared online, they are no longer protected by privacy laws; and, Judge Ma. Cecilia Austria Case in which a personal social media account of a government official was deemed inappropriate, setting precedent for government personnel to not use social media where publicly accessible.

media.[91] Policies of relevance can be found in the Constitution, Revised Penal Code[92], and Rules of Court and Civil Code.[93]

The Anti-Video and Voyeurism Act of 2009, meanwhile, was passed quickly after internet access began to grow in the state.[94] The law seeks to limit and discriminate the posting and sharing of unconsensually filmed sexual acts by penalty of incarceration.[95] During the last five years, this act has been important in enforcing disinformation campaigns which involve deep fakes and fabricated images of influential people engaging in taboo behavior.

Further, the Data Privacy Act of 2012 is one of two cybersecurity bills, which acknowledge the need for individual protections online. This law specifically protects personal information and systems for communication in the private and public spheres. It requires that institutions "preserve the integrity and confidentiality of all personal data...and sensitive personal information." The act also helped to further clarify and distinguish data processing protocols pertaining to personal information. Notably, this extends to data harvesting and states that individual consent is mandated. This act came before the Philippines was a testing site for Cambridge Analytica in 2016 but highlights the vulnerability of internet access preceding digital literacy.
In a similar thread, the Cybercrime Prevention Act of 2012 intended to prevent cybercrimes related to sex, sexuality, and specifically predatorial behavior as well as cybercrime specifications on identity theft and spamming.[96] The law made cybercrimes punishable by up to a 12-year sentence.

---

[91] Karen Brutas, "Social Media and Privacy: The Philippine Experience," Foundation for Media Alternatives, accessed May 11, 2020, https://www.fma.ph/resources/resources-gender-ict/social-media-and-privacy-the-philippine-experience/.

[92] Title 13 of the Revised Penal Code of the Philippines criminalizes libel and the spread or promotion of slander. Due to mistreatment of journalists, the UN Commission on Human Rights ruled that the criminalization of libel ultimately contradicts ICCPR Article 19, stating: "Defamations laws should not… stifle freedom of expression…[and] Penal defamation laws should include defense of truth."

[93] "The 1987 Constitution of the Republic of the Philippines - Article III"; "Philippines: New 'Cybercrime' Law Will Harm Free Speech," Human Rights Watch, September 28, 2012, https://www.hrw.org/news/2012/09/28/philippines-new-cybercrime-law-will-harm-free-speech.

[94] Disini & Disini Law Office, "The Anti-Photo and Video Voyeurism Act of 2009: A Primer," Data Privacy Philippines, March 2, 2016, https://privacy.com.ph/articles/the-anti-photo-and-video-voyeurism-act-of-2009-a-primer/.

[95] This law is relevant because some advocates have unsuccessfully attempted to use the law as a precedent for banning violent or graphic content images shared on social media in the Philippines.

[96] "Philippine Cybercrime Law Takes Effect amid Protests," BBC News, October 3, 2012, https://www.bbc.com/news/technology-19810474.

Perhaps most relevant to this case study is the Free Public Internet Access Program, signed into law in August 2017, formally known as the Republic Act 10929. The act mandates free internet access in public spaces including government offices, state universities, hospitals, and parks. According to current data from the Department of Information and Communications Technology, the policy led to the installation of more than three times as many access points in the succeeding two years after the law was passed juxtaposed to the previous 32 months under the former *Free Public Wi-Fi* project, a notable spike from 807 installations to 2,708 in two-thirds the time.[97]

Most recently, in January 2018, Facebook and the current Duterte administration partnered to build an undersea cable system that connects Philippine internet systems to the rest of Asia and the US—a project aimed to improve the Philippines' poor internet infrastructure and better enabling internet penetration.[98]

PART TWO: Politically Driven Violence is Systemic

Since overthrowing a violent, class polarizing regime in 1987, incidents of conflict have centered on periods surrounding elections.

The "ora de peligro" (the hour of danger) is when most political interventions occur. This period for every Philippine election since 1988 has occurred as much as two weeks preceding a vote, with violent incidents growing until election day. During this window, candidates have engaged in vote buying and coercion, where voters have experienced threats of violence to vote for a candidate or to abstain from voting. [99]

Electoral violence and political candidates have historically been involved in corruption, including gang wars over illegal profiteering from the Philippine government. Some analysts state the Philippines is a weak state because of the prevalence of violence

[97] "Duterte Signs Law on Free Internet Access," CNN, August 3, 2017, https://cnnphilippines.com/news/2017/08/03/Duterte-signs-law-on-free-internet-access.html; Leo Carlo Liay, "Philippines: Capturing the Broadband Satellite Opportunity," The Asia Foundation, August 28, 2019, https://asiafoundation.org/2019/08/28/philippines-capturing-the-broadband-satellite-opportunity/; "Free Wi-Fi for All Hotspots," Government of the Philippines, May 7, 2020, http://freepublicwifi.gov.ph/livehotspots/.
[98] ABS-CBN News, "Gov't Partners with Facebook for High Speed Broadband," ABS-CBN News, November 15, 2017, https://news.abs-cbn.com/news/11/15/17/govt-partners-with-facebook-for-high-speed-broadband.
[99] Patrick Patino and Djorina Velasco, "Election Violence in the Philippines" (Manila, Philippines: Friedrich Ebert Stiftung, 2004).

connected to local politics: a key component of a weak state, is its "monopoly over the legitimate use of violence."[100]

Campaign tactics have traditionally included electoral violence such as verbal threats or attacks on candidates and opponent supporters; attacks on rallies and infiltrating communication mediums; kidnappings; destruction of campaign materials and tactics of intimidation such as open carrying in opponent supported communities.

The Philippines has faced a series of conflicts that, scholars have argued, were driven by politics—with communal violence observed to emerge preceding and immediately after local and national elections. Electoral violence takes a range of forms including abductions, targeted attacks in private or public settings, and physical violence in some cases, resulting in physical injury and fatalities. Dangerous speech has historically been present in many of these conflicts, with candidates often exacerbating current issues related to disparities including class, public health, education, and culture.

Social Media as a Weapon

One of the primary factors contributing to electoral violence is the participation of the national police and military. Public trust in the police and military have historically been very low, with high perceptions of corruption and links to criminal networks. Another factor is nepotism. Family members have repeatedly run and won elections locally and nationally, which amplified small- and large-scale tensions centered on corruption and a lack of just opportunities in the state.

The weaponization of social media has also potentially exacerbated conflict in the Philippines, aggravating pre-existing tensions on political corruption through a campaign centering on drug use and crime.

Social media amplifies dangerous speech in a targeted and personalized nature, which has been proven to exploit fundamental cognitive processes and influence key stakeholders with greater speed and efficiency than other mediums of media.[101] In general, social media can lead to rapid polarization among identities, as individual's opinions are reinforced through product algorithms and data-driven advertisements.

---

[100] Patino and Velasco.

[101] Norman Vasu et al., "Fake News: National Security in the Post-Truth Era" (Singapore: Nanyang Technological University Singapore, January 2018).

At the same time, rumors gain credibility which create further divisions among vulnerable populations. And in cases such as Myanmar and the Philippines, the drivers of conflict are swifter and, perhaps, faster now than past approaches of information-driven threats of violence, such as newspapers, live television, or even radio.[102]

Among conflict typologies vulnerable to social media misuse examined in the Mercy Corps report, The Weaponization of Social Media, is "Electoral Violence," defined as "violence used by political operatives and supporters to achieve their victory." Using Kenya as an example, the report states how social media platforms "amplify political competition between constituencies divided along ideological or identity lines, in some cases inciting hatred and violence."[103]

Exposing Vulnerability: Elections and Social Media

The role of social media in the 2016 election of Duterte and dangerous speech paired with Facebook intersects with a range of socio-historic predispositions to communal violence.

The impact of trolls in the Philippines, where Facebook's prominence and the state's vulnerability to polarization and communal conflict, were critical factors in Duterte's campaign tactics which enabled his team to influence dialogue online.[104]Although trolls and bots consist of approximately five percent of global Facebook users, the proportion is significantly higher in the Philippines.[105]

And since Duterte assumed power in July 2016, the Philippines has seen thousands of Filipinos murdered, dozens of politicians and journalists arrested, and a rise in dangerous speech made by the president.[106]

---

[102] Sian Herbert, "Conflict Analysis: Topic Guide" (Birmingham, UK: GSDRC University of Birmingham, May 2017).

[103] "The Weaponization of Social Media: How Social Media Can Spark Violence and What Can Be Done about It" (Portland, Oregon: Mercy Corps, Do No Digital Harm, and Adapt Peacebuilding, November 2019). Kenya background and social media: communal conflict exacerbated by social media usage as a medium for rapidly and efficiently spreading dangerous speech.

[104] Henry Mihm, Ines Oulamine, and Fiona Singer, "The Philippines Deserves More From Facebook," Lawfare, October 27, 2019, https://www.lawfareblog.com/philippines-deserves-more-facebook.

[105] "An Update on How We Are Doing At Enforcing Our Community Standards," About Facebook, May 23, 2019, https://about.fb.com/news/2019/05/enforcing-our-community-standards-3/.

[106] "Philippines: Duterte's 'Drug War' Claims 12,000+ Lives," Human Rights Watch, January 18, 2018, https://www.hrw.org/news/2018/01/18/philippines-dutertes-drug-war-claims-12000-lives.

As evidenced in the introduction, electoral violence is not uncommon in the Philippines, where political trends of polarization are systemic. For example, paying more than 500 individuals to form "troll armies" and create fake multiple social media accounts for government officials and candidates to vilify their opponents is a modernized approach to similar approaches historically.[107]

2.1 Facebook is the Internet

The rapid rise to Filipinos' usage of the platform began in October 2013 when the networking site rolled out free access to Philippine smartphone users without data charges through an initiative known as "Free Facebook." The plan was to partner with a local carrier to offer a portal of free, basic internet services. Since then, the Philippines' internet use and accessibility contribute to international trends of global expansion.[108] As Figure 1.2 demonstrates, Facebook use far surpasses any social media platform in the country.

Facebook first attempted a similar project in 2010: "Facebook Zero," an initiative aimed to offer free mobile access to users with a reduced-feature version.[109] The project's objective was in part to increase its international presence in over 45 countries particularly in Asia, Latin America, and Africa.

In 2013, the social media giant rolled out a program Free Basics, in partnership with six other companies to bring affordable access to selected internet services to underdeveloped countries - and was most successful in the Philippines.
To enact such wide scale product availability, Facebook partnered with *Globe*, then the smaller of the telecommunications duopoly, trailing by 20 percentage points after *Smart,* the main player. After the partnership, however, *Globe's* numbers ballooned— enough to overtake its primary competition in just 15 months. In a 2014 conference in

---

[107] Shanshank Bengali and Evan Halper, "Troll Armies, a Growth Industry in the Philippines, May Soon Be Coming to an Election near You," Los Angeles Times, November 19, 2019, https://www.latimes.com/politics/story/2019-11-19/troll-armies-routine-in-philippine-politics-coming-here-next.

[108] Bianca Consunji, "Facebook Rolls Out Zero Data Charge Access in the Philippines," Mashable, November 1, 2013, https://mashable.com/2013/11/01/facebook-philippines/; "Globe Telecom to Offer Free Facebook Access for Its 36 Million Subscribers," Inquirer.net, October 23, 2013, https://technology.inquirer.net/30705/globe-telecom-to-offer-free-facebook-access-for-its-36-million-subscribers; Jessi Hempel, "Inside Facebook's Ambitious Plan to Connect the Whole World," Wired, January 20, 2016, https://www.wired.com/2016/01/facebook-zuckerberg-internet-org/.

[109] Ben Parr, "Facebook Launches New Mobile Site with No Data Fees," Mashable, May 18, 2010, https://mashable.com/2010/05/18/0-facebook-mobile-site/.

Barcelona, Facebook founder Mark Zuckerberg acknowledged the rapid assimilation saying, "What we've seen in the Philippines is... a home run."[110]
As Facebook revolutionized how information was spread in the Philippines, a country called the social media capital of the world by a global media agency consecutively in recent years, it was not as successful in other countries in the region, such as India, mainly due to pushback by local governments, telecommunication companies, and human rights organizations.[111]

2.2 Drivers of Dangerous Online Speech

    i.   Personalization of Facebook posts, videos, and live streaming improves target isolation

Facebook uses algorithms which cater to each individual user's interests. Each user curates their Facebook account through posts they like, comment on, or links they open through the platform. The technology of the medium takes the personalization further through social media metadata, aiding content producers to efficiently and accurately tailor and target individuals prioritized and visible content. This process of data mining and curating has made Facebook and other social media platforms commercially viable.[112]

At the same time, the process lends itself to rapid spread of dangerous speech, influencing attitudes, beliefs, and potentially behaviors connected to the inciting of violence. And in the case of the Philippines, it allowed trolls to promote false content and ultimately lead a narrative that promoted Duterte's ideologies and political campaigns.

Further, the Philippines is vulnerable to conflict as evidenced by its long history of political corruption, socioeconomic inequities, and violent conflicts especially inflated during elections or time periods of political transition. Around the world, weak states with rapidly growing social media penetration are experiencing spikes in conflict as the internet and political infrastructure capable of mitigating harm is slower to form.

---

[110] "Facebook CEO Mark Zuckerberg: Philippines a Successful Test Bed for Internet.Org Initiative with Globe Telecom Partnership," Globe, February 25, 2014, http://www1.globe.com.ph/press-room/philippines-a-successful-test-bed-for-internet-org-initiative.
[111] Kemp, "Digital 2019: The World."
[112] "The Weaponization of Social Media: How Social Media Can Spark Violence and What Can Be Done about It."

2.3 A New Era of Dangerous Speech: Duterte leverages Facebook to win the elections

With pre-existing vulnerabilities to disinformation and a new expansion of Facebook use, 2016 then-presidential candidate Duterte capitalized on this phenomenon, mounting his campaign mainly on social media. His campaign stood on two primary platforms: (1) anti-illegal substances, namely "shabu" (the local name for methamphetamines), and (2) anti-elite.

In a state with more than 20 percent living in extreme poverty, Duterte's stance on increasing wealth redistribution and lowering drug use resonated with Filipinos. Many believed that his campaign against illegal drugs was addressing a severe public health emergency, exacerbated by crime, lack of education, and poverty.

And in May 2016, Duterte had an unprecedented win with nearly 30 percent of the 54 million votes for the presidential election.[113]Since then, an internal Facebook report has labeled Duterte the "undisputed king of Facebook conversations."[114] What Duterte lacked in political machinery—politicians who are often affiliated to influential clans and or organized parties—he was able to make up for in strategy.

Duterte, for his campaign, spent the least among his rivals on political campaigns overall and rather, his team invested their limited funds on the internet as opposed to the traditional television advertising which can amount to millions of pesos by the minute. Duterte's campaign manager, Nic Gabunado, admitted in 2018 that the campaign spent around US$200,000 on social media and hired 400 to 500 trolls to "amplify" Duterte's ideas, according to a University of Oxford study.[115]The same study also found that the month before the election, 64 percent of all Philippine election posts on Facebook mentioned Duterte.

---

[113] "Philippines Election: Maverick Rodrigo Duterte Wins Presidency," BBC News, May 10, 2016, https://www.bbc.com/news/world-asia-36253612.

[114] Etter, "What Happens When the Government Uses Facebook as a Weapon?"

[115] Samantha Bradshaw and Philip N. Howard, "Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation" (University of Oxford: Computational Propaganda Research Project, December 2017).

Figure 2.1

By leaning on the way Facebook was intended to work for individuals, trolls were able to manipulate dialogue and plant disinformation. At the time of the election (May 2016), Facebook's news feed algorithm method—of pushing to the top public content that generated the most shares, likes, and comments—was essential for managers to exploit public Facebook accounts and Twitter. They did this by hiring trolls who ran dozens of fake accounts and who were skilled at curating personal pages by amplifying posts concerning Duterte by using expletives or shocking rumors about his opponents and anyone who publicly criticized the president.[116]

And, among the most Duterte-pushing accounts, only 26 are responsible for the majority of all viral posts containing dangerous speech and graphic images of Duterte's campaigns, according to Philippine online media company Rappler.[117] Collectively the accounts' posts engaged with more than four million Filipino Facebook users with each account followed by the others—and only the others alone.

---

[116] Rob Price, "Facebook Banned 2.2 Billion Fake Accounts in the First 3 Months of This Year. That's Almost Equal to the Number of Real People Who Use It.," Business Insider, May 23, 2019, https://www.businessinsider.com/facebook-bans-2-billion-fake-accounts-q1-2019-2019-5.

[117] Maria A. Ressa, "Propaganda War: Weaponizing the Internet," Rappler, October 3, 2016, http://www.rappler.com/nation/148007-propaganda-war-weaponizing-internet.

In a comment made by Ellen Tordesillas, the President of Vera Files, a Facebook fact-checking partner in the Philippines, she stated that the "majority" of fake posts her organization detected during the Philippines election season were churned out by pro-administration Facebook pages.[118]

One alarming example was a proven fake sex tape of an opposing party member and Duterte critic, Philippine Senator Leila de Lima. The fabricated post went viral and has had long lasting impacts on Filipinos' trust in de Lima as a political figure. Facebook reported it is in the process of improving its content moderation in the Philippines, particularly when the tools on the platform could do harm.[119]

After Duterte won the election, the administration demonstrated further uses for social media. The president banned independent media organizations from covering his live inauguration and aired the event on Facebook Live. With the help of the state's television, all cameras live-streamed through the social networking site.[120]

2.4 Techniques and Tactics for Political Manipulation

During the Summer of 2016, Duterte encouraged supporters to conduct mass killings using statements such as "Do it yourself if you have a gun, you have my support," and "Shoot him and I will give you a medal" in explicit reference to people he considers "drug personalities." Prior to his campaign, Duterte was the Mayor of Davao for 22 years.[121] His primary focus was similar to those in his platform and centered on illegal shabu use. During his time as mayor, the local police force was connected to 1,000 deaths questioned by human rights defenders internationally. Duterte has publicly bragged about committing violence in various instances, saying he personally killed

---

[118] Davey Alba, "How Duterte Used Facebook to Fuel The Philippine Drug War," BuzzFeed News, September 4, 2018, https://www.buzzfeednews.com/article/daveyalba/facebook-philippines-dutertes-drug-war.

[119] Alba.

[120] Raissa Robles, "Journalists Banned from Live Reporting Duterte's Swearing-in Ceremony as President," South China Morning Post, June 28, 2016, https://www.scmp.com/news/asia/southeast-asia/article/1982677/journalists-banned-live-reporting-dutertes-swearing.

[121] "Kill Drug Dealers and I'll Give You a Medal, Says Philippines President," The Guardian, June 5, 2016, https://www.theguardian.com/world/2016/jun/05/kill-drug-dealers-medal-philippines-president-rodrigo-duterte; Emily Rauhala, "Before Duterte Was the Philippines' President, He Was 'the Death Squad Mayor,'" Washington Post, September 28, 2016, https://www.washingtonpost.com/world/asia_pacific/before-duterte-was-the-philippines-president-he-was-the-death-squad-mayor/2016/09/28/f1d1ccc4-800b-11e6-ad0e-ab0d12c779b1_story.html.

drug users and peddlers in Davao and had every intention to do the same as the chief executive.[122]

Political disinformation campaigns generally center key design strategies and implementation tactics. In the case of the Philippines, most political candidates employ campaign management or public relations strategists to use market-driven techniques to engage local audiences with consistent messaging. These messages, according to a Mercy Corps Report, are successful in "weaponizing vernaculars to maximize the read of social media posts."

Campaign strategist Gabunado, for instance, crafted a wide-reaching plan to hire anonymous people to troll social media accounts and promote the campaign's strategies. This included educating volunteers on the brand of the campaign, which highlighted drug use as a danger to the infrastructure of the state, and those who use or distribute it as dangerous. In turn, volunteers utilized similar tactics as hired account holders, or trolls, to increase Duterte-centered conversations on Facebook and Twitter.

Figure 2.4 is an example of a viral meme, shared 88,000 times on Facebook with more than 600 comments. English translation: *Fellow DDS (Diehard Duterte supports), let us hold on to Father Digong (nickname of Duterte), no leaving each other behind until the end. 1 share = 1 support.*

Figure 2.2

---

[122] Megan Specia, "Duterte, Philippine President, Boasts He Killed Someone as a Teenager," The New York Times, November 10, 2017, https://www.nytimes.com/2017/11/10/world/asia/philippines-president-duterte-killed.html.

Political candidates also tap Social Media Influencers (people defined as having 50,000 or more social media followers and those who specialize in popular culture) at times. Duterte for his part, appointed internet-famed lifestyle blogger Mocha Uson as the PCOO assistant secretary in charge of social media, receiving backlash from the public.[123] Uson, who lacks government experience or a background relating to managing a communications team as those holding her position previously have, had more than five million Facebook followers at the time of her appointment, most of whom are Filipino. Her understanding, therefore, of the media was valued by the current administration.

In 2017, the administration appointed another social media blogger, Lorraine Badoy, who, like Uson, calls herself a "Diehard Duterte Supporter."[124] Badoy is the current PCOO undersecretary. Uson and Badoy have been accused by civil society and international human rights organizations for attacking the president's critics online through Facebook livestream videos and public comments.[125]

Figure 2.3

[123] Rappler.com, "FALSE: Mocha Uson Says She 'Does Not Spread Fake News,'" Rappler, October 6, 2018, http://www.rappler.com/newsbreak/fact-check/213594-false-mocha-uson-does-not-spread-fake-news; Maan Macapagal, "PNP: 23,000 Killings under Investigation in 2 Years," ABS-CBN News, June 13, 2018, https://news.abs-cbn.com/news/06/13/18/pnp-23000-killings-under-investigation-in-2-years.

[124] Pia Ranada, "Duterte Online Defender Appointed Social Welfare Assistant Secretary," Rappler, February 13, 2017, http://www.rappler.com/nation/161304-lorraine-badoy-assistant-secretary-dswd. The term, "Diehard Duterte Supporter," has the same acronym, "DDS", as the Davao Death Squad: a vigilante group made up of anonymous members who claim to have murdered thousands of alleged drug users, distributors or associated persons between 2016 and 2019.

[125] Patty Pasion, "DSWD Exec Tells EU: 'Mag-Child Porn Muna Kayo,'" Rappler, March 28, 2017, http://www.rappler.com/nation/165388-dswd-undersecretary-badoy-eu-child-pornography.

Figure 2.4
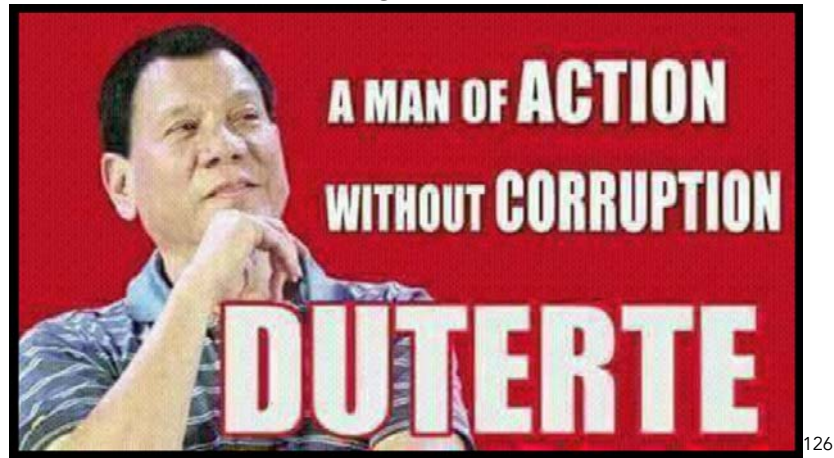
A strategy candidates also use is amplifying their personal messages through fake accounts by Contracting employees who remain anonymous to localize key campaign narratives. Although many trolls or bots use scripts and copy and paste text into comment threads of videos or posts, the people hired to work on Duterte's campaign used a specific formula in order to escape algorithms aimed to make the platform safer. They never paste text into comments; instead, they post graphic images, buried within popular threads, making these harder for Facebook moderators to find and take down. The troll farms engaged at a grassroots level by having back-and-forth dialogue promoting disinformation with real account holders.

These specific troll techniques, according to Malou Tiquia, chief executive of the Manila political strategy consultancy Publicus Asia, "will be used more and more." She said, "When Facebook said it won't ban political ads, that was already a signal to everyone that anything goes.[127]

---

[126] Popular meme shared more than 200,000 times. Rappler reports originated from Uson.

[127] Bengali and Halper, "Troll Armies, a Growth Industry in the Philippines, May Soon Be Coming to an Election near You."

Figure 2.5

Frequency of Tweets Mentioning Each Presidential Candidate
Nov 22 - Nov 25, 2015
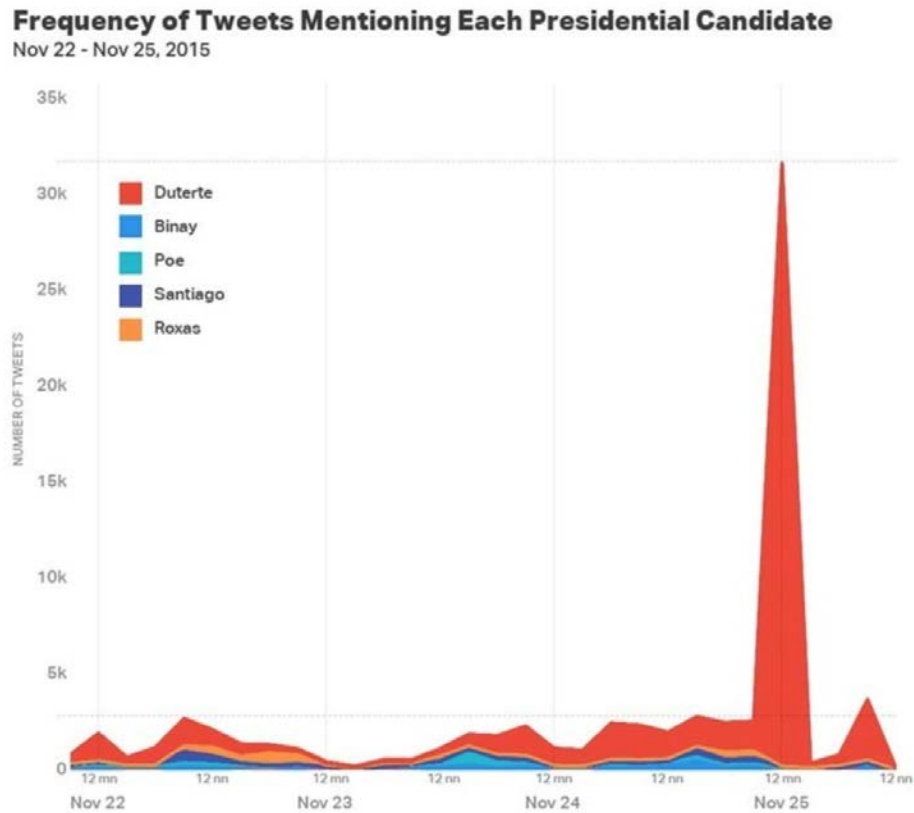
2.5 Response of Key Stakeholders

A. Public opinion

The majority of Filipinos do not believe dangerous speech concerning alleged drug users, distributors, or associated people is a problem The latest national poll conducted by the Social Weather Station revealed his popularity hitting a new high for the last three months of 2019—with an "excellent" net satisfaction rating and 80 percent approval rating.[128]According to the same source, approval rating was down by 10 percent from the start of the administration in 2016. Duterte's supporters trust him to do what is necessary for the development of the Philippines, and when asked, 65 percent approve of his tactics on the war on drugs. Commenters often call Duterte their "King." On Facebook, the majority of comments centering Duterte's campaign on drugs are positive towards the situation and declare their support for the president.

---

[128] Andreo Calonzo, "Duterte's Popularity Hits New High, Pollster SWS Says," Bloomberg, January 21, 2020, https://www.bloomberg.com/news/articles/2020-01-22/duterte-s-popularity-hits-new-high-pollster-sws-says.

B. Media

Major media outlets have been targeted by Duterte's administration for publishing stories that were critical of their policies and subsequent implementation. Tensions between the media and the administration have been historically high in the Philippines, with this administration building on these trends. Legally, the nation does not have regulations that prevent journalists from publishing and have freedom of speech clauses in their constitution, bill of rights, and recent legislation.[129] But in February 2020, the government moved to shut down ABS-CBN, the country's biggest media conglomerate, which employs over 10,000 people in what was seen by critics as one among many attempts by Duterte's administration to silence their detractors who question the drug war.[130]

Award-winning journalist Maria Ressa, who founded Rappler, was, meanwhile, arrested multiple times in 2019. The year before that, the government threatened the closure of her entire organization and banned a Rappler reporter from entering the entire Malacanang complex (where the President resides and where press conferences are held).[131]

C. Multilateral institutions

Responding to concerns on Duterte's alleged human rights violations, the United Nations' Human Rights Council in July 2019 supported a resolution brought forward by Iceland to investigate these killings, expressing concern over the "staggering" amount of deaths during Duterte's three-year campaign[132] It mandates Michelle Bachelet, head of the world body's Human Rights Council, to look into the deaths in Duterte's war on

[129] Carlos H. Conde, "A New Weapon Against Press Freedom in the Philippines," Human Rights Watch, December 6, 2018, https://www.hrw.org/news/2018/12/06/new-weapon-against-press-freedom-philippines.

[130] Jason Gutierrez, "Philippines Moves to Shut ABS-CBN, Its Leading Broadcast Network," The New York Times, February 10, 2020, https://www.nytimes.com/2020/02/10/world/asia/philippines-abs-cbn-franchise.html.

[131] Phelim Kine, "Philippine Government Targets 'Rappler' for Closure," Human Rights Watch, January 15, 2018, https://www.hrw.org/news/2018/01/15/philippine-government-targets-rappler-closure; "Rappler Reporter Now Banned from Entire Malacañang Complex," Rappler, February 22, 2018, http://www.rappler.com/nation/196601-reporter-banned-entire-malacanang-complex.

[132] Martin Petty and Paul Tait, "'Not a Chance' - Philippine Minister Says No Access for U.N. Drugs War Probe," Reuters, September 11, 2019, https://www.reuters.com/article/us-philippines-drugs-idUSKCN1VW0JL.

drugs.[133] It also calls for a comprehensive written report into the human rights situation in the country. The resolution marks the first time the 47-member council has been asked to address the ongoing violations. It is also the first time an investigation will be conducted in a country where there is no civil war.[134]

Philippine officials condemned the measure saying it is "one-sided" and "hypocritical." The Foreign Affairs Secretary warned of "consequences" after the vote.[135]The Office of the President goes further, calling it "grotesquely one-sided, outrageously narrow, and maliciously partisan."[136]Duterte himself, three days after the resolution was filed, stated in a speech before his department officials, "What is the problem of Iceland? Ice only. That is your problem." He added that the Nordic country has no understanding of the socioeconomic and political problems in the Philippines.[137]

D. International community

Organizations outside the Philippines have also raised their concerns regarding Duterte's war on drugs. Thus far, Duterte has received international condemnation by the European Commission, Human Rights Watch, Amnesty International, among others.[138]

Governments have also offered help as well. Japanese Prime Minister Shinzō Abe in 2017 extended an $800-million financial assistance for Philippine drug rehabilitation

---

[133] Petty and Tait.

[134] Jeoffrey Maitem, "UN Probe into Duterte's Philipines Drugs War Welcomed by Activists - but Manila Calls Move an 'Insult,'" South China Morning Post, July 13, 2019, https://www.scmp.com/news/asia/southeast-asia/article/3018408/dutertes-drug-war-human-rights-activists-welcome-uns.

[135] John Reed, "Rodrigo Duterte Suspends Aid Deals with Countries Backing UN Probe," The Financial Times, September 20, 2019, https://www.ft.com/content/d0af30f8-db8f-11e9-8f9b-77216ebe1f17.

[136] Petty and Tait, "'Not a Chance' - Philippine Minister Says No Access for U.N. Drugs War Probe."

[137] Maitem, "UN Probe into Duterte's Philipines Drugs War Welcomed by Activists - but Manila Calls Move an 'Insult.'"

[138] Enrico Dela Cruz, "Philippines Decries European Parliament's 'Interference,'" Reuters, April 20, 2018, https://www.reuters.com/article/us-philippines-drugs-eu-idUSKBN1HR05L; "World Report 2019: Rights Trends in Philippines," Human Rights Watch, December 28, 2018, https://www.hrw.org/world-report/2019/country-chapters/philippines; "Stop Extrajudicial Executions in the Philippines," Amnesty International, 2020, https://www.amnesty.org/en/get-involved/take-action/stop-extrajudicial-executions-in-the-philippines/.

centers although made no mention of deaths resulting from the drug war.[139] Indonesia also expressed apprehension, staying, "Shoot on sight policy leads to abuse of power. We still believe in the presumption of innocence." Yet in September 2016, the head of Indonesia's National Anti-Narcotics Agency said the agency was contemplating on copying the Philippines' hardline tactics against drug traffickers, a comment, quickly downplayed after.[140]

Taiwan and China adopted a different tone, however, welcoming Duterte's plan to declare a war against criminality and illegal drugs.[141]Chinese officials said, they "fully understand that the Philippine government under the leadership of H. E. President Rodrigo Duterte has taken it as a top priority to crack down drug-related crimes." The statement, again, made no reference to extrajudicial killings, and called illegal drugs the "common enemy of mankind".[142]

E. Facebook

Headquartered in Manila, Facebook's Philippine branch is growing.[143]And, following advocacy from Rappler reporters in November 2016, Facebook's cyber security department confirmed that the campaign strategies of Duterte's team violated the platform's user policies and removed hundreds of pages and a dozen troll run accounts which, through their six months on the site, reached 3.6 million users.[144]

In statements released on how Facebook aims to prevent future violations and the spread of disinformation and dangerous speech, they will focus on increasing digital

[139] Phelim Kine, "Philippine President Rodrigo Duterte's 'War on Drugs,'" Human Rights Watch, September 7, 2017, https://www.hrw.org/news/2017/09/07/philippine-president-rodrigo-dutertes-war-drugs.

[140] Krithika Varagur, "Indonesia Borrows Philippines Rhetoric to Revamp Drug War," Voice of America, August 11, 2017, https://www.voanews.com/east-asia-pacific/indonesia-borrows-philippines-rhetoric-revamp-drug-war.

[141] Patricia Lourdes Viray, "Taiwan Confident of Stronger Ties with Philippines under Duterte," PhilStar Global, June 14, 2016, https://www.philstar.com/headlines/2016/06/14/1592910/taiwan-confident-stronger-ties-philippines-under-duterte.

[142] ABS-CBN News, "China Supports Duterte's War on Drugs," ABS-CBN News, July 19, 2016, https://news.abs-cbn.com/news/07/19/16/china-supports-dutertes-war-on-drugs.

[143] Miguel R. Camus, "Facebook Expands Presence in PH, Opens New Office," Inquirer.net, May 31, 2018, https://business.inquirer.net/251707/facebook-expands-presence-ph-opens-new-office.

[144] Gelo Gonzales, "Facebook Takes down Fake Account Network of Duterte Campaign Social Media Manager," Rappler, April 7, 2019, https://www.rappler.com/technology/news/226932-facebook-takes-down-fake-account-network-duterte-campaign-social-media-manager-march-2019.

literacy and removing fake accounts through increased attention to patterns, such as spikes in specific names, new accounts made, ages, or post-dates–a tactic which in six months of implementation, doubled fake account removals.[145]

Still, as the internet and social media grow exceptionally in the Asia-Pacific region, the company states that many clients continue to unintentionally misuse the platform, as well as social media products generally.

Policy statements and transparency reports by Facebook, including the annual Community Standards Enforcement Report,[146] demonstrate attempts to navigate this challenging and fast-paced situation, aggravated by current events and pre-existing norms and precedents daily.[147] In February 2020, Facebook global content policy Vice President released a white page report, *Charting a Way Forward*, detailing the company's cyber security strategy specific to "harmful speech" and content moderation. In the report she states that the company will strive to increase content monitoring through global expansions of context specific literacy tools to empower local populations advancing their bottom line and ensuring appropriate use of the medium.[148]

The program, "We think Digital" is a localized digital literacy program aimed to achieve these objectives. In the Philippines this is known as, "Digital Tayo" (In English:

---

[145] Adam Mosseri, "Working to Stop Misinformation and False News," About Facebook, April 7, 2017, https://www.facebook.com/facebookmedia/blog/working-to-stop-misinformation-and-false-news; "Removing Coordinated Inauthentic Behavior From the Philippines," About Facebook, March 29, 2019, https://about.fb.com/news/2019/03/cib-from-the-philippines/; "Community Standards Enforcement," About Facebook, November 2019, https://transparency.facebook.com/community-standards-enforcement#hate-speech; Rizal Raoul Reyes, "Facebook Rolls out Digital Tayo Program," BusinessMirror, May 11, 2019, https://businessmirror.com.ph/2019/05/11/facebook-rolls-out-digital-tayo-program/; Vanessa Romo and Amy Held, "Facebook Removed Nearly 3.4 Billion Fake Accounts In 6 Months," NPR, May 23, 2019, https://www.npr.org/2019/05/23/726353723/facebook-removed-nearly-3-2-billion-fake-accounts-in-last-six-months.

[146] "Community Standards Enforcement."

[147] In the Philippines, the spread of disinformation on COVID-19 has led to policies criminalizing those who defy health expertise. See more information here Gustavo Guerra, "Philippines: Criminal Charges Filed over Misinformation on Coronavirus," web page, Library of Congress - Global Legal Monitor, March 2, 2020, //www.loc.gov/law/foreign-news/article/philippines-criminal-charges-filed-over-misinformation-on-coronavirus/; Fatima Gaw, "Digital Disinformation Is as Potent as a Virus during a Pandemic," Rappler, March 24, 2020, http://www.rappler.com/technology/features/255224-digital-disinformation-fake-news-coronavirus.

[148] Monika Bickert, "Charting a Way Forward: Online Content Regulation" (Facebook, February 2020).

We are Digital).[149] The program launched in January 2020 and educates users on how to identify fake pages or posts by bots and paid trolls. Still, Facebook's cyber security team has stated that if people still succumb to fake posts, the company cannot be held liable.

Further, according to Facebook's community standards report, approximately five percent of their global monthly account users were made up of fake accounts, and in the Philippines this number is reportedly higher. Based on data disclosed in their quarterly financial reports, in the last quarter of 2019, there were over 1.7 billion identified and removed fake accounts globally.[150] Their data in, however, are estimates and according to a spokesperson, are not "designed to match population or census estimates" and that, Facebook's data is targeted at informing businesses of advertising reach, stating the estimator was initially designed to "give marketers an estimate of how to plan their campaigns on Facebook."[151]

Although Facebook has made public commitments to removing fake accounts and increasing digital literacy, the market focus leaves inherent gaps in the social, cultural, and historic dynamics in which it is used. This leaves clear potential for dangerous abuse of the platform, leading to the vulnerable spread of disinformation and dangerous speech.

PART THREE: The War on Drugs, Propelled by Duterte's use of Dangerous Speech

3.1 Background on the War on Drugs

Since beginning his campaign in 2016, Duterte has used dangerous speech to promote his war on drugs. He has constantly aggravated controversy and justified inhumane treatment and violence against Filipinos through dehumanizing tactics, reducing people he defined as associated with shabu as "slaves of drug peddlers," "beyond redemption," "do-nothings," "dead-living Filipinos," "rapists who kill their own mother," or vermin who destroy the foundation and survival of the state. He has

---

[149] "We Think Digital," We Think Digital - Facebook, 2020, https://wethinkdigital.fb.com/; "Digital Tayo," Facebook Digital Tayo, accessed May 12, 2020, https://wethinkdigital.fb.com/ph/en-us/resources/.

[150] "Community Standards Enforcement."

[151] Vernise L. Tantuco, "Did You Know That Some PH Cities Have More Facebook Users than Actual Population?," Rappler, February 1, 2019, http://www.rappler.com/newsbreak/iq/215159-list-philippines-cities-more-facebook-users-than-population.

portrayed alleged people who use or distribute drugs as dangerous people who incite violence against their families, communities, and the nation.[152]

By August 1, 2016, a month into Duterte's war on drugs, there were already 900 drug-related killings, 700 anti-illegal drug operations, 700 arrests, and hundreds of thousands of voluntary surrenders all over the country. While the Philippine government has declared at least 6,600 deaths in the anti-drug campaign, human rights groups and international agencies estimate the state-sponsored killings to be an upwards of 29,000.[153]

An investigation led by Columbia University's Toni Stabile Center for Investigative Journalism revealed that the figures declared by the government were a gross underestimation of the extent of drug-related killings in the Philippines. The data the center collected proved that the killings of drug suspects, by both police and unidentified shooters, have been purposely excluded from official counts.[154]

Stabile's findings concluded that the disparity goes beyond reporting incidents of violence, but more on how the country's war on drugs was being carried out.[155]As several media reports have exposed, many Filipinos are being denied due process, targeted for assassination, and offered little to no chance of rehabilitation.

On September 30, 2016, three months after Duterte's inauguration, the president held a televised press conference wherein he stated: "Hitler killed 3 million Jews… There are 3 million drug users. There are. I would be happy to slaughter them."

---

[152] "Bawal Ang Pasaway: PNP Chief Dela Rosa: Nanlaban Ang Mahihirap Na Pusher Kaya Napapatay," YouTube, August 30, 2016, https://www.youtube.com/watch?v=Jnr5su9ezX4; Marlon Ramos, "'Shabu Addicts beyond Redemption,'" July 21, 2016, https://newsinfo.inquirer.net/797735/shabu-addicts-beyond-redemption; Sheila Coronel, David Mora, and Mariel Padilla, "The Uncounted Dead of Duterte's Drug War," The Atlantic, accessed May 12, 2020, https://www.theatlantic.com/international/archive/2019/08/philippines-dead-rodrigo-duterte-drug-war/595978/; Dahlia Simangan, "Is the Philippine 'War on Drugs' an Act of Genocide?," *Journal of Genocide Research* 20, no. 1 (January 2, 2018): 68–89, https://doi.org/10.1080/14623528.2017.1379939.

[153] "MAP, CHARTS: The Death Toll of the War on Drugs," ABS-CBN News, March 9, 2020, http://newsprod.abs-cbnnews.com/specials/map-charts-the-death-toll-of-the-war-on-drugs; Stephanie Nebehay, "Philippines Faces Call for U.N Investigation into War on Drugs Killings," Reuters, July 5, 2019, https://www.reuters.com/article/us-philippines-rights-un-idUSKCN1TZ22M.

[154] "Philippine Drug War Casualties," data.world, 2019, https://data.world/stabile-center/ph-drug-war.

[155] Coronel, Mora, and Padilla, "The Uncounted Dead of Duterte's Drug War."

Duterte's speech and excerpted quote is an example of dangerous speech because it is 1) explicit in his intent to kill and, 2) it implied his assumed impunity and justification in doing so.

3.2 Analysis of Dangerous Speech and the War on Drugs

Social/Historical Context

Duterte was the mayor of Davao City for 22 years, a period of violence and hard crackdowns against alleged criminals, drug users, and distributors. Elected in 2016, Duterte is the 16th president of the Philippines.

In his speeches, Duterte often identifies two primary target groups: drug users, specifically of shabu, and who he deems are "drug personalities," which includes people sympathetic to drug use (e.g. journalists, human rights defenders, some politicians and their constituents and select police officers).[156] In his September speech, he referred to a running hit list as he falsely reminded followers of an inflated number of drug users: connecting the use of drugs to a justified murder and implying impunity for those who killed.

Duterte has flagged shabu users as the primary threat to the state, a substance he believes to be most used by lower income Filipinos.[157] On the other hand, the current administration has minimized those who use cocaine or heroin, drugs, the president has stated, are mostly used by the rich. This stance has been backed by the PNP and is also supported by the Armed Forces of the Philippines.

Drug use in the Philippines is significant, with 2.3 percent or 1.8 million of the adult population currently using an illegal substance.[158] When Duterte took office, some 1 percent of adults reported using shabu. Hundreds of thousands of youth were also reported to be either using or selling illegal drugs. While these figures are significant, Duterte's claim in September 2016 that more than three million people chronically use

---

[156] Yvonne T Chua and Maria Diosa Labiste, "President Rodrigo Duterte's Dangerous Speech in the Philippines," Dangerous Speech Project, October 2017, https://dangerousspeech.org/wp-content/uploads/2019/10/Philippines_Chua-and-Labiste_Final.pdf.

[157] "TIMELINE: The PNP's Use of the Term 'Deaths under Investigation,'" Rappler, accessed May 12, 2020, http://www.rappler.com/newsbreak/iq/165534-timeline-philippines-pnp-deaths-under-investigation.

[158] "Operations as of October 28, 2017," Republic of the Philippines Philippine Information Agency, October 28, 2017, https://pia.gov.ph/socialcards/cards/142.

drugs is false. Data from the WHO show the state's drug use has been on a steady decline for the past decade, with more than 6.8 million users in 2004.

Supporters of Duterte's war on drugs believe the campaign is reducing decades of political corruption and underground criminal networks. In a primarily Catholic country, illegal drug use is shamed, and the eradication of drug use is seen as a positive for the nation.

The figure Duterte may have referred to the 3.8 million adults who, as of 2019, the Philippine Dangerous Drug Board data stated have tried an illegal substance once in their lifetimes.[159]

Within the first three months of his presidency, July to September 2016, more than one million people turned themselves into Philippine National Police (PNP) for either using or distributing shabu and nearly half a million more were arrested and held in overwhelmed and overcrowded jails.[160]

The primary speech analyzed demonstrates efficient strategies Duterte used through social media to gain popular support for his campaign on drugs. The timeframe of study is the immediate five months after Duterte's inauguration in July 2016, covering the months of July to November 2016. During this period the campaign's killings both peaked and gained media, international and sectorial criticism. In the middle of the period, Duterte shifted his strategy to center on impunity and elite protections to those conducting violence from justifications and convincing statements on the eradication of alleged drug users or distributors.

Duterte's statements are explicit: he has stated repeatedly that he wants drug personalities dead rather than wounded, arrested, or even rehabilitated. He has instructed specific groups to kill and stated that if they do, he will grant them impunity. A recent report by Amnesty International states his speech on September 30, 2016 can be interpreted as a direct order to police and vigilantes, to kill people he has defined as suspects.

The President justifies drug-related killings by declaring those who use or distribute as immoral people, often citing unfounded claims that drug users, peddlers, or

---

[159] Patrick Winn, "In the Philippines, Populist Rage Is Directed at Addicts, Not Immigrants," PRI The World, June 29, 2016, https://www.pri.org/stories/2016-06-29/philippines-populist-rage-directed-addicts-not-immigrants. Marijuana use is cited as the most common illegal substance tried.
[160] Winn.

sympathizers are rapists and murderers. This justification is further entrenched by a multifaceted reinforcement by the PNP, social media trolls, and followers that drug-related killings are justified, stating drug personalities fired the first shots, thus reducing the narrative to self-defense. Finally, he again argues that shabu use is a threat and eliminating users and their allies will save the poorest communities in the country.

Duterte, in a number of his speeches, has referred to alleged drug users as people who have lost capability of functioning and at best, burdensome to their communities.[161]

Many of Duterte's statements condemning drug users and threatening violence, some of the most controversial statements have been to the PNP and the military, those with the authority to enforce the war on drugs.[162] Thus amplifying dangerous speech where the message is explicitly directed at an in-group or key population able to incite action. In the Dangerous Speech report on the Philippines, a highlighted excerpt from Duterte's August 17, 2016 speech directed at the national police, he stated:

> "We have seen our country devastated by drugs. And it has not only affected millions, but a lot of them are no longer viable as human beings on this planet.…And what is really very unsettling is that a year or more of shabu use would shrink the brain of a person, and therefore he is no longer viable for rehabilitation. Let us now say that there are about 400 drug addicts no longer eligible for rehabilitation for they are really crazy and out of their senses and no longer have the cognitive value of their person or their talent. So, what do we do with it? We've about 300 dead living Filipinos?"

Soon thereafter on, August 26, 2016, Duterte made the following comments in an address to the military:

> "Sabi ko nga, (As I said) crime against humanity? In the first place, I'd like to be frank with you. Are they humans? What is your definition of a human being? Tell me. Human rights. Use it properly in the right context, if you have the brains. Now, if your gray matter between the ears is melting, I cannot help you, kung ganoon ang tingin mo. (if you see it that way.)"[163]

---

[161] Derogatory statements excerpted include: "liki," "wala na sa landas," and "bangag."

[162] "Philippines: The Police's Murderous War on the Poor," Amnesty International, January 31, 2017, https://www.amnesty.org/en/latest/news/2017/01/philippines-the-police-murderous-war-on-the-poor/.

[163] Chua and Labiste, "President Rodrigo Duterte's Dangerous Speech in the Philippines."

Duterte's Speeches Begin to More Explicitly Sanction Violence Against "Drug Personalities"

In the first three months of Duterte's war on drugs, his speeches were primarily targeted at the PNP and military. His tone and language center the wrongdoings of alleged "drug personalities." In these speeches he repeats a justified narrative of killing specified targets. But by October, when the war attracted international attention and human rights advocates began criticizing the campaign and unwarranted and potentially illegal killings under the ICCPR simultaneous to civil society filing lawsuits against the administration's actions, Duterte's speeches shifted. He began reinforcing his promise of impunity for those who act. He told police that he would grant them amnesty and personally defend them in court. At the same time, data on police and military initiated killings began to decline. Juxtaposed to the shift in speech in October 2016, vigilante led murders began to climb.

Among these groups is the Davao Death Squad, or DDS, an anonymous group of devout Duterte followers who are believed to be responsible for the murder of thousands of Filipinos and drug-related arrests of even thousands more in Davao, and soon, when he became president, nationwide.[164] The way the group operates is mainly through a public Facebook page where they solicit anonymous clues for targets.[165] They then publish the mugshots of people arrested, murdered, or physically harmed as a result of the war on drugs. Whether or not the people posted on their page are truly victims of the group is unknown. However, with only 804 followers, the page sees many more shares and engagement than the number lets on. Graphic images of crime scenes color the page, and at the time of writing this report, hundreds of images remain visible on the platform.

But the DDS is not alone; dozens of vigilante groups and individuals have been tracked across the country by journalists and organizations including Human Rights Watch and Amnesty International.[166]

---

[164] Felipe Villamor, "Ex-Officer in Philippines Says He Led Death Squad at Duterte's Behest," The New York Times, February 20, 2017, https://www.nytimes.com/2017/02/20/world/asia/rodrigo-duterte-philippines-death-squad.html.

[165] "Davao Death Squad," Facebook, January 24, 2016, https://www.facebook.com/DavaoDeathSquad82.

[166] "Philippines," January 31, 2017.

Police officials and military members have been found to use Facebook Messenger for coordinating official business.[167] From plotting raids to discussing cases, several police officials have publicly and anonymously flagged Facebook as a primary medium used in the field.

Another group of anonymous vigilantes claiming to carry out drug-related killings are The Davao Boys. Reputable researchers have tracked their Facebook posts charting their self-promotion and involvement in the homicides and successfully crowdsourced suspects in identifying connected members to several drug-related killings where witnesses were present.

Fiercely loyal, many of Duterte's most outspoken Facebook followers hail from Davao. Among those who are police officials, many followed the former Mayor to Manila upon election, posting on Facebook about their transfer and staunch support of the President-elect.[168]

The Importance of Media in Disseminating Duterte's Dangerous Speech

At the start of Duterte's administration, Facebook users spent an average of three hours 16 minutes a day online.[169] But between July and December, this number grew to an average of four hours and 12 minutes. In the same timeframe, keywords of "Duterte," "Shabu," "Davao Boys," "DDS," and "Justice," dominated up to 65 percent of Facebook posts and video taglines.

Unknowingly or not, when Duterte made his infamous September speech, he ignited a rapid-fire discussion online. Critics, supporters, and those in-between went to the platform, fanning the flames of a dangerous period overwhelmed by communal executions. While many expressed fears for their lives and that of their loved ones, others became emboldened—and Facebook noted a crowdsourced increase in user violation reporting.

---

[167] Clare Baldwin and Rew R. C. Marshall, "How a Secretive Police Squad Racked up Kills in Duterte's Drug War," Reuters, December 19, 2017, http://www.reuters.com/investigates/special-report/philippines-drugs-squad/.

[168] Baldwin and Marshall.

[169] Simon Kemp, "Digital 2016: The Philippines" (Hootsuite and WeAreSocial, 2016), https://datareportal.com/reports/digital-2016-philippines.

By October of 2016, journalists called on Facebook to increase their content moderation and investigate the spread of disinformation and inflammatory comments, videos and posts.[170]

Through a deep analysis of videos posted from the PCOO to YouTube and Facebook, this report found that between July and November of 2016, views, comments and shares of speeches increased each month with subsequent journalistic re-sharing adding to the magnitude of scale.

The President's speeches, which are public, have repeatedly gone viral on Facebook, beginning in 2015 to present day in 2020. These speeches form the base of his following.[171]

The administration has a sustained trend of excluding journalists, television broadcasters, or radio hosts from recording public facing press conferences.[172] For public addresses, Facebook is the primary medium where messages are live streamed or hosted on YouTube. Through social media's wide rate of consumption, the speeches receive anywhere from hundreds of thousands to millions of views. In particular, his regular addresses to the nation receive a rounded average of 678,000 views per video, with some having more than 2.5 million views.[173]

Many of these statements have been turned around and brought back to the public. Some believe his communications team converts dangerous speech into memes, replicating pop culture and reinforcing a particular narrative in an accessible and relatable way to target populations, including young people.[174]

---

[170] Dana Priest, "'Seeded in Social Media': Jailed Philippine Journalist Says Facebook Is Partly Responsible for Her Predicament," Washington Post, February 25, 2019, https://www.washingtonpost.com/technology/2019/02/25/seeded-social-media-jailed-philippine-journalist-says-facebook-is-partly-responsible-her-predicament/.

[171] "Made in the Philippines: Fake Social Media Accounts and Trolls with Global Ambitions," The Washington Post, July 25, 2019, https://www.washingtonpost.com/.

[172] Manuel Mogato, "Philippines Complains Facebook Fact-Checkers Are Biased," Reuters, April 16, 2018, https://www.reuters.com/article/us-philippines-facebook-idUSKBN1HN1EN.

[173] Russia Today, "'They Want Me to Fight China. It's Gonna Be a Massacre!,'" YouTube, May 22, 2017, https://www.youtube.com/watch?v=rHjlCmdyesY.

[174] Amber Newman, "Troll Palayan: Clara Balaguer on Design, Decolonization, and Trolling Duterte," Walker, May 11, 2018, https://walkerart.org/magazine/troll-palayan-clara-balaguer-on-design-decolonization-trolling-duterte.

Nationally and locally in Davao, Duterte is among the most widely reported on public officials in the history of private media in the Philippines. With 20 television networks and hundreds of radio stations, including Radio Television Malacañang (RTVM), People's Television Network; Bureau of Broadcast Services, Rappler, and the Intercontinental Broadcasting Corp, nearly every media company in the nation actively covers the administration. Many of which post transcripts and follow-up analysis of his speeches.

Duterte's September 2016 statement was directed to the public from his headquartered office in Manila and has since received more than four million views.[175]In terms of location, Duterte delivers his speeches in three primary settings: his office, at authority headquarters for the PNP, and at the AFP (both of which are key to the war on drugs, having the authority to enforce the administration's campaigns).

All presidential speeches not live broadcasted by television stations are posted to the website of the PCOO where the video, transcripts and sometimes audio clips are freely available. The Presidential Broadcast Staff manages the network, RTVM, through which they post most of the administration's videos on YouTube, who also hosts most media produced, and on Facebook. Each platform has its own following, on YouTube RTVM has 161 thousand subscribers, with RTVM's Facebook page having 240 thousand followers.[176]

3.3. Dangerous Speech Timeline and Data Analysis

It is important to note the data discrepancies between domestic and international analyses when considering the connection between social media use and dangerous speech because current data show rampant corruption in the Philippines. In the annual global index on corruption from the organization Transparency, the Southeast Asian country ranked 113 out of 180 states globally.[177]

---

[175] Eli Moskowitz, "Philippines' Top Drug Enforcer: Violent Fear Tactics Are Failing," Organized Crime and Corruption Reporting Project, February 10, 2020, https://www.occrp.org/en/daily/11588-philippines-top-drug-enforcer-violent-fear-tactics-are-failing.

[176] "RTVMalacanang - YouTube," Youtube, July 25, 2008, https://www.youtube.com/user/RTVMalacanang/videos; "Radio Television Malacañang - RTVM," Facebook, accessed May 12, 2020, https://www.facebook.com/pg/rtvmalacanang/videos/?ref=page_internal.

[177] "Philippines," Transparency International, 2019, https://www.transparency.org/country/PHL.

The primary discrepancies in confirming the true number of extrajudicial killings is a result of three primary factors:

1. Prolific cover ups by the PNP and the military, including records of planting shabu on a victim following an assault.[178]
2. Beliefs in an untraced disposal of bodies juxtaposed with record numbers of missing persons, many of whom family members believe were vulnerable to penalization from the war on drugs and,[179]
3. An 85 percent surge in shooting victims seen in hospitals between June 2016 and January 2017.[180]

Figure 3.1[181]

---
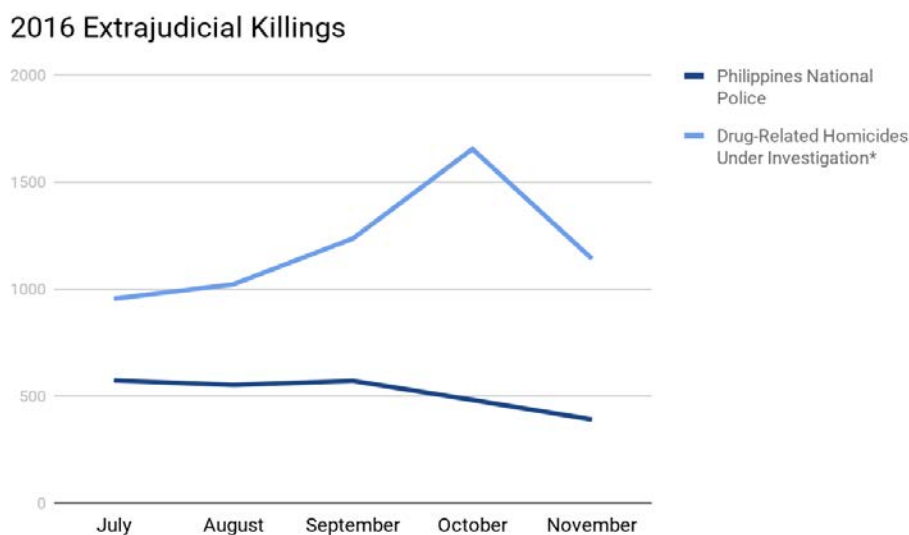
[178] Coronel, Mora, and Padilla, "The Uncounted Dead of Duterte's Drug War."

[179] "Philippine Police 'dumping Bodies' of Drug War Victims," July 28, 2017, https://www.aljazeera.com/news/2017/07/philippine-police-dumping-bodies-drug-war-victims-170728034001676.html.

[180] Clare Baldwin and rew R. C. Marshall, "Philippine Police Use Hospitals to Hide Drug War Killings," Reuters, June 29, 2017, http://www.reuters.com/investigates/special-report/philippines-duterta-doa/.

[181] "Annual Report 2016" (Manila, Philippines: Philippine National Police, 2016); "Registered Deaths in the Philippines, 2018," Republic of the Philippines Philippine Statistics Authority, February 5, 2020, https://psa.gov.ph/vital-statistics/id/159965. The PNP data was taken directly from their 2016 annual report while those under investigation are an inconclusive number garnered for this report. Specifically, the number was found through tracing the logs of local mortician reports, the Philippines Statistics Authority, and the extensive qualitative data compiled by Human Rights Watch, Rappler, and Amnesty International, isolating this explicit period and comparing these findings against case filings among the PNP for homicide investigations. We then used a statistical analysis program in STATA to combine the respective data, find their average and incorporate a 1.659% margin of error.

Figure 3.1 represents drug-related fatalities in the Philippines between a targeted period of two months preceding President Duterte's statement likening himself to Hitler and two months thereafter: July 2016 to November 2016, also the first five months of Duterte's presidency.

A 2017 report released by social media analysis company Graphika, which helped Facebook identify content violations on the site, affirmed the posts removed glorified Duterte's war on drugs.[182]

In January 2019, the social media company again deleted more than 200 Facebook pages and 73 accounts, including 29 Instagram accounts in what it called "one of the most significant takedowns in the world."[183] One of the pages, which belonged to Twinmark Media Enterprises, was followed by 43 million accounts, majority from the Filipino community. Facebook informed the government about the move. Twinmark is the company behind Uson's page, a loyal Duterte follower.[184]

Head of Facebook's Cybersecurity Policy, Nathaniel Gleicher, stated that what alerted the company was Twinmark's consistent violation of Facebook's spam policies and its manipulation of public information through coordinated inauthentic behavior, use of fake accounts, and selling administration access to Facebook pages in order to increase profit.

"We have notified the Philippine government already regarding it, but we don't recommend what the government should do." Gleicher said. "We're more focused on removing the violating behavior from our platform." [185]

---

[182] Melanie Smith, "Archives: Facebook Finds 'Coordinated and Inauthentic Behavior' In the Philippines; Suspends a Set of Pro Government Pages Ahead of May Elections," Medium, March 29, 2019, https://medium.com/graphika-team/archives-facebook-finds-coordinated-and-inauthentic-behavior-in-the-philippines-suspends-a-set-d02f41f527df.

[183] Kaela Malig, "Facebook Takes down 220 Pages, 73 Accounts, 29 Instagram Accounts," GMA News Online, January 11, 2019, https://www.gmanetwork.com/news/money/companies/681063/facebook-takes-down-220-pages-73-accounts-29-instagram-accounts/story/.

[184] Paige Occeñola and Gelo Gonzales, "PH Company Banned by Facebook Spread Lies, Used Fake Accounts," Rappler, April 16, 2019, https://www.rappler.com/technology/social-media/220741-facebook-remove-trending-news-portal-twinmark-media-enterprises.

[185] "Banning Twinmark Media Enterprises in the Philippines from Facebook," About Facebook, January 11, 2019, https://about.fb.com/news/2019/01/banning-twinmark-media-enterprises/.

Facebook began its investigations as early as November 2018, when it first identified the group selling access to their Facebook pages.[186]

Figure 3.2

2016
Facebook Use: Timeline of Focus

JULY — **MONTH ONE**

Facebook use similar to May and June: 3 hours and 16 minutes/day. Duterte's livestream speeches on Facebook average 89,134 views per video.

AUGUST — **MONTH TWO**

Facebook use increases, with people spending more time online per day. Video and images posted increase. Reputable images of extrajudicial killing crime scenes go viral .
Duterte's livestream address to the nation average of 156,141 views.

SEPTEMBER — **MONTH THREE**

Facebook use again increases.
Dutertre livestreams speech containing dangerous speech promoting the massacre of 3 million people.
Across media platforms, millions more were watching - Facebook alone had 132,562 views).

OCTOBER — **MONTH FOUR**

Facebook use averages to more than 4 hours a day.
Duterte's livestream speeches average 192,492 views.
Protests and international pressure increases.
Facebook begins removing posts.

NOVEMBER — **MONTH FIVE**

Facebook use again averages more than 4 hours a day.
Duterte's livestream speeches average 234,244 views.
Facebook begins plans with Duterte to increase internet access in rural parts of the state.

Figure 3.4

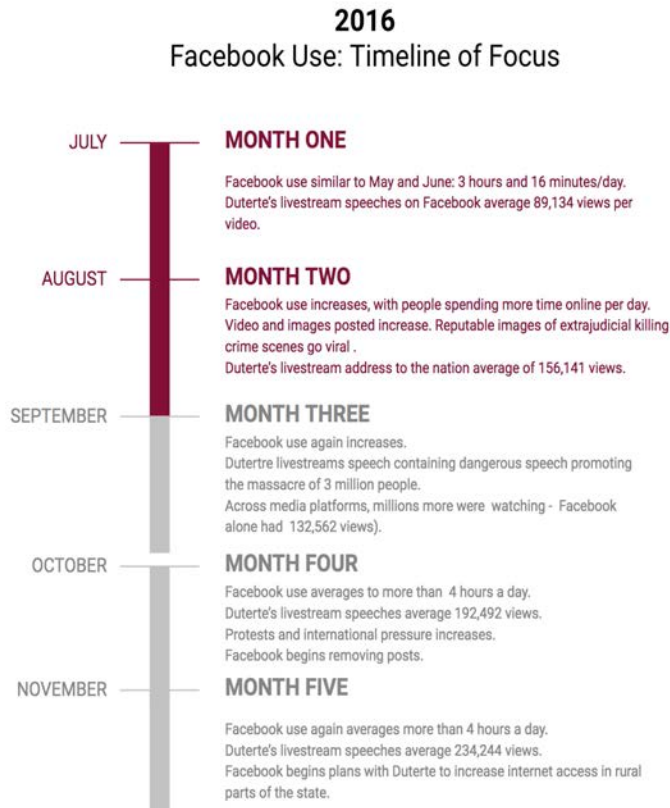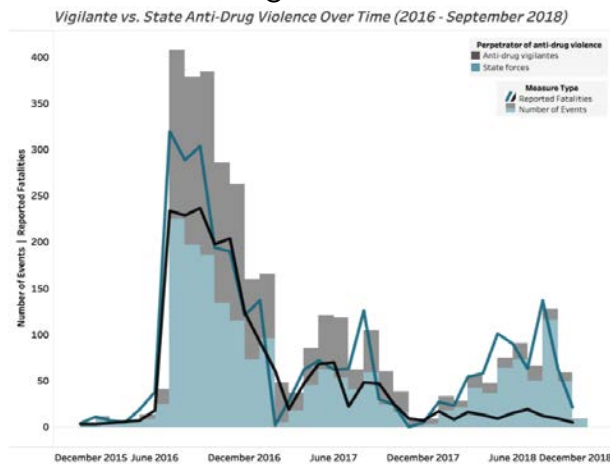Vigilante vs. State Anti-Drug Violence Over Time (2016 - September 2018)

---

[186] Malig, "Facebook Takes down 220 Pages, 73 Accounts, 29 Instagram Accounts."

3.4. Responses

Supporters

Duterte's supportive base stems from long standing histories of distrust in career politicians and associated crime networks. Although Duterte was Mayor for 22 years, supporters do not see him as a traditional politician because he does not come from a lineage of policymakers like many of his predecessors. His proven track record on cracking down on petty crimes and low-level illegal drug offenses demonstrate to those who have been affected both by small scale crime and drug addiction in their community is appealing.

Running on a campaign that emphasized the corruption of previous leaders and that of his competition attracted constituents weary of another term under more traditional political leadership.

Allies of the president's ideological stance on shabu and the extermination of alleged users and distributors are far reaching, ranging from government officials, police heads, and justices. Among these are Police Chief Ronald del Rosa, whom in August 2016, made a comment to people who had turned themselves in for activities condemned by the drug war operations that his police would support them in killing drug lords or committing arson against their properties. Additionally, he has agreed with Duterte that low-income people are a "natural target" of the drug war because many are, "drug crazed." The class polarization is emphasized because, del Rosa says poor people resist arrest more often than rich people

Other allies include elected persons such as the Speaker of the House who states the war on drugs is just and that a shoot-to-kill order would be more appropriate than the status quo. And at the time of the 2017 UN Universal Periodic Review where the Philippines faced scrutiny over human rights violations, then Foreign Secretary Alan Peter Cayetano was tasked with defending the war on drugs at the UN General Assembly.

In the Department of Justice, which houses the National Prosecution Service, Justice Secretary Vitaliano Aguirre has defended the president's ideologies in many accounts. Namely concerning international human rights law. Aguirre echoed the president's statements that alleged drug users have "no humanity," further implying that drug users are inhuman and therefore undeserving of humane protection before the law.

Duterte's online followers are among the most vocal and active Facebook accounts in the country, with many actively posting, tweeting, and making videos hailing their support of Duterte's success in eliminating alleged drug-related problems. Uson, now an employee of the president, grew her following through Duterte supportive videos where she crafted accessible and short videos explaining the war on drugs in addition to regularly re-posting the president's speeches and photo montages from the "battle ground." Her posts have had over 3 million views, 110,345 likes and 119,942 shares. Some followers have crowdsourced video projects where they report on the war on drugs. One Facebook video series tracks the fatalities from the war as "points" for Duterte's success and popularity as president.

Counterspeech

The scope and scale of Duterte's war on drugs, his manipulation of Facebook, recruitment, and purchase of anonymous people to carry out his missions have brought forth many supporters and critics alike.

Internationally, fact finding reports have been published by the United Nations, Human Rights Watch, and Amnesty International. The former Obama Administration has released a series of condemning statements, calling on world leaders to invoke sanctions and publicly condemn Duterte's war on drugs.

The Roman Catholic Church has also condemned the Philippine president, a country where nearly 80 percent are Catholic, and his subsequent condoning of mass violence during his campaign. During a speech, then Presidential candidate Duterte falsely promoted an endorsement from Pope Francis. The Roman Catholic Church quickly released statements that the Pope does not endorse Duterte and, further, that the Church does not align with his tactics, policies, nor his leadership.

International and domestic lawyers, politicians, celebrities, journalists, and civil society have also rejected Duterte's propaganda and labeled him as a corrupt leader with disregard for human rights and criticized his administration for promoting disinformation. The Philippines Commission of Human Rights, dozens of local universities and student groups have taken to social media, namely Facebook, to criticize the administration and defend those targeted by the war on drugs.

Both Amnesty International and Human Rights Watch have released recommendations to international allies, aiming to increase pressure against Duterte's war on drugs, and for Duterte and government agencies themselves to mitigate harm and cease the war.

And as quickly as Duterte took office, protesters flooded the streets and individuals have taken it upon themselves to research and follow the war on drugs in their communities. Photojournalists work at night, tracking Facebook comments, publishing kill lists and attempting to document the facts behind drug-related killings.

Journalists who advocated for increased democratic principles and policies have extensively documented the lies, propaganda, and systematic violence incited by Duterte and its inherent link to Facebook. Rappler journalists, a leader in reporting on the administration, have a particular stake in the matter: their primary homepage is hosted on Facebook. Rappler founder Ressa helped Facebook advance in the Philippines at its earliest stages, believing the platform would increase dialogue, education, and opportunities.

In March 2016, Ressa hosted a Facebook sponsored livestream interview with presidential candidates, an idea she thought would even out the political landscape which favors wealthy candidates who can pay for swaths of advertisements. Duterte was the only candidate to show and the interview received 1.9 million views.

Ressa identified him as a threat to the safety and health of many in the country and has been vocal to Facebook officials, domestic and international policy managers about the dangers Facebook poses to the Philippines. Rappler has thoroughly documented the prevalence and use of trolls as opposed to bots and was the first organization to document these strategies by the administration.

Facebook officials have made numerous attempts to mitigate the danger the platform enables. In March 2019, Facebook again removed a cluster of hundreds of posts, accounts and group pages linked to Duterte's campaign and social media manager, Gabunada.[187] Together, as many as 4 million of the accounts followed at least one of the removed pages whose content a Facebook report stated promoted disinformation and were against the platform's user guidelines. Each month, Facebook reports removing hundreds of posts containing graphic or violent content in the Philippines.

Rappler and Reuters reports document how Facebook has failed to moderate dangerous content and that in spite of these efforts, violent images, livestreams and posts explicitly calling for murder, assault, extinction and rape to stay on the platform

---

[187] Price, "Facebook Banned 2.2 Billion Fake Accounts in the First 3 Months of This Year. That's Almost Equal to the Number of Real People Who Use It."

for months – at the time of this report, thousands of graphic posts in the Philippines remain intact.[188]

PART FOUR: Conclusion

Globally, addressing dangerous speech at the intersection of evolving social media trends and penetration demands a comprehensive approach, layered with multi-stakeholder perspectives and obligations.

The roots of national conflict are inherently context specific, but in many instances, tensions are rapidly exacerbated by the fast-paced nature of social media. As it expands, the potential for wide scale violence is eminent. Community enacted and authority enabled violence is ubiquitous to times of political transition in the Philippines. And, as the leading nation of social media users in time spent online, companies need to reconcile their role in these preexisting dynamics, especially with a product which available data has demonstrated, amplifies polarization and hostilities of an us-versus-them paradigm.

In the Philippines, Facebook is integrated into the landscape of daily life. It is not enough to end trolling solely given the complexity of the situation. To do so would not counter the spread of disinformation and hostilities personally promoted by influential people, including celebrities, authorities such as police officers, judges, and military officials or elected officials. Emphasizing the need for digital literacy is critical in the state. And, social media companies should follow Facebook's initiative in requiring users to receive regular training integrated into their use of the product.

Ultimately, social media companies need to be more aggressive in countering dangerous speech on their sites. In March 2019, Facebook announced a policy defining its parameters on banning posts promoting or making related statements to white supremacy.[189] In a similar thread, this caliber of content regulation is needed in the Philippines. If Facebook demonstrated this frame of leadership in the Philippines, other emergent social media markets in the region may follow suit and increase their bottom line for a safer, more connected product and ultimately, shift its role in

---

[188] Mogato, "Philippines Complains Facebook Fact-Checkers Are Biased"; Mihm, Oulamine, and Singer, "The Philippines Deserves More From Facebook."

[189] Sasha Ingber, "Facebook Bans White Nationalism And Separatism Content From Its Platforms," NPR, March 27, 2019, https://www.npr.org/2019/03/27/707258353/facebook-bans-white-nationalism-and-separatism-content-from-its-platforms.

amplifying state-sponsored disinformation and dangerous speech among millions internationally.

NIGERIA
CASE STUDY

With nearly 200 million citizens, Nigeria is the most populous country in Africa as well as the continent's largest economy. In the last few decades, the country has become a technological hub—one of the top ten oil-exporting countries worldwide—and an increasingly important diplomatic and cultural leader in the sub-region. Nigeria is also among the most diverse countries in the world. While English is the national language—one of many legacies of the colonial era that only ended in 1960—there are around 500 languages spoken within the country including Hausa, Yoruba and Igbo. Most of the population in the north are Muslim and in the south are Christian.

Despite many gains, 40 percent of Nigerians live on less than a dollar a day.[190] In 2018, Nigeria overtook India to become the country with the most people living in extreme poverty globally, and currently ranks 158 out of 189 countries and territories on the UN Human Development Index. Nigeria has experienced numerous security challenges since independence including a devastating civil war in the late 1960s, communal conflict in the central region (the middle belt) in the early 2000s, and the Boko Haram insurgency in the north east, which continues to this day. Hundreds of Nigerians are killed or displaced by identity-based conflict each year.

While the roots of identity-based conflict are based in historical experiences and local conditions, recent technological advancements are changing the way Nigerians understand current security threats. Social media has greatly accelerated the reach and speed of messages shared online. In some cases, this has improved government accountability and democratized access to information. In others, dangerous speech has exacerbated pre-existing religious and ethnic tensions and increased the likelihood that violence will occur.

This case study investigates the role that social media platforms play in shaping identity-based violence in Nigeria. Specifically, this report provides an examination and analysis of the actors, online platforms, and practices that may increase the risk or severity of communal violence. Part one of this report provides general information about Nigeria's social media landscape. Part two looks at how WhatsApp and Facebook, specifically, have been weaponized in Nigeria, focusing on how the dissemination of misinformation and dangerous messages intensify or exacerbate conflict. Part three uses the dangerous speech framework to illustrate how dangerous speech has impacted communal violence between farmers and herders in central and eastern Nigeria. Part four concludes the Nigeria section of this report.

---

[190] "Forty Percent of Nigerians Live below the Poverty Line: Report," Al Jazeera, May 4, 2020, https://www.aljazeera.com/news/2020/05/forty-percent-nigerians-live-poverty-line-report-200504141558636.html.
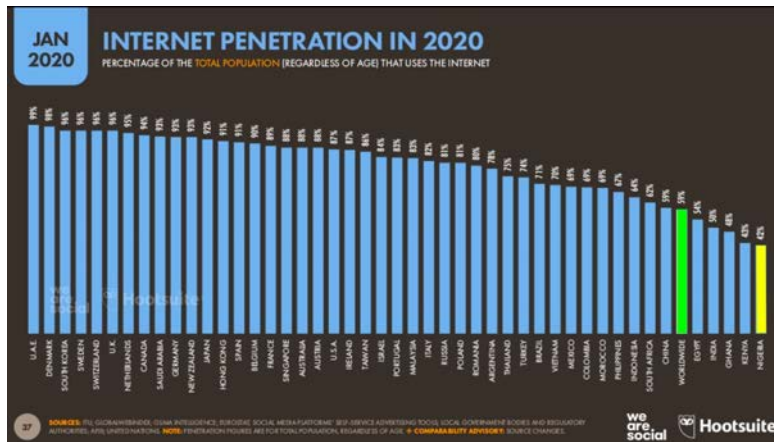
PART ONE: Nigeria's Media Landscape

Newspapers and radio have historically dominated Nigeria's media landscape. Under military rule that lasted until 1999, these traditional media sources operated under strict government control with little freedom to publish or share dissenting voices. Today, Nigeria has a vibrant media landscape where commentators openly criticize the government and their policies.[191] At the same time, the country ranks 115 of 180 countries on the Reporters Without Borders World Press Freedom Index.[192] Journalists are commonly surveilled, attacked, intentionally stymied, and even arbitrarily arrested or killed when covering politics, terrorism, or investigating corruption.

The internet was introduced in Nigeria in the year 2000 but still remains largely inaccessible to the general public. It is only within the last decade that internet use expanded, and social media emerged, though penetration remains relatively low. Currently, Nigeria ranks fourth in the world—after India, China, and Pakistan—for having the greatest number of people living without internet access. Nearly 120million Nigerians do not use the internet, or 58 percent of the total population. One reason for this is the high cost of mobile data, more expensive in Africa than anywhere in the world, is prohibitively expensive for much of Nigeria's population. In contrast, 83 percent of Nigeria's population owns a mobile phone, which remains the most utilized medium of interpersonal communication.[193]

---

[191] "Freedom of the Press 2017 - Nigeria," Freedom Hoouse, April 28, 2017, Nigeria, https://www.ecoi.net/de/dokument/1418604.html.

[192] "Nigeria : Climate of Permanent Violence," Reporters Without Borders, accessed May 12, 2020, https://rsf.org/en/nigeria.

[193] mon Kemp, "Digital 2020: The World," We Are Social, January 30, 2020, https://wearesocial.com/blog/2020/01/digital-2020-3-8-billion-people-use-social-media; Iamosi Ekenimoh, "Overpriced: Why Nigerians Pay so Much for Data," Stears Business, April 11, 2019, /article/overpriced-why-nigerians-pay-so-much-for-data; mon Kemp, "Digital 2020: Nigeria," We Are Social, January 30, 2020, https://datareportal.com/reports/digital-2020-nigeria.

INTERNET PENETRATION IN 2020
PERCENTAGE OF THE TOTAL POPULATION (REGARDLESS OF AGE) THAT USES THE INTERNET

Social media penetration varies widely by region on the African continent. As a percentage of the population, North Africa leads the continent with 39 percent of the region's population using social media, followed by Southern Africa (36 percent), West Africa (13 percent), East Africa (8 percent), and Middle Africa (6 percent). While less than half of Nigeria's total population are able to access the internet, an even smaller percentage is active on social media. Nigeria has 27 million social media users, or roughly 13 percent of the population, and is equivalent to the average regional percentage of social media users in West Africa. Social media platforms are accessed almost exclusively via smart phones, 99 percent of the time.[194]

1.1 USERS & REACH

Social media use in Nigeria is still very much an urban phenomenon with far less access available in rural areas. Trends in targeted social media advertising reveal that users are overwhelmingly young and more likely to be male. According to We Are Social data, Facebook reports 67 percent of social media advertising targets users ages 18 to 34. In addition, 62 percent of its advertising audience are reported as male and 38 percent are female. Although there are more men than women using Facebook, data show women "like" and share more posts than men do. Information about the specific content of the posts women like and share is unavailable. However, because dangerous speech is often spread and promoted on this platform, female users may be key actors to involve in media literacy efforts given their disproportionately high levels of engagement.[195]

---

[194] Kemp, "Digital 2020," January 30, 2020.

[195] Kemp.

According to experts, Facebook and WhatsApp users in urban settings are likely to write and share messages in English. However, the further away from cities, the more users share messages in local languages. In the north especially, many messages are shared in Hausa, the regional lingua-franca. Voice notes are also popular among populations who lack formal education or do not communicate in a written language.

While social media penetration in Nigeria is relatively low, posts shared on social media platforms may have a disproportionate reach when combined with more ubiquitous traditional media platforms such as radio, print media, and television. This happens in a number of ways. Radio or talk show TV hosts may read social media posts on air, broadcasting messages to thousands of listeners who otherwise would not have access to them. Radio remains the key source of information for Nigerians, especially in rural areas that lack internet access. Forty-four percent of respondents in one recent study reported they listened to the radio daily.[196]

Additionally, print media sources have been found to reproduce false content shared on social media. Lack of adequate resources and comprehensive training as well as a repressive political environment contribute to this problem. As a result, rumors and other false or misleading messages that originated on social media reach audiences who otherwise would not access online platforms. In addition to spreading dangerous speech and misinformation, this practice undermines journalistic credibility, further exacerbating the vacuum of credible information, and increasing mistrust of the mainstream media's institutional knowledge and authority.[197]

---

[196] Idayat Hassan and Jamie Hitchen, "Driving Division? Disinformation and the New Media Landscape in Nigeria" (Abuja, Nigeria: Centre for Democracy & Development, February 2020), 15.
[197] Hassan and Hitchen, 16.

1.2 KEY PLATFORMS

The two most common social media platforms in Nigeria, determined by the percentage of users aged 16-64 who report using each platform within the last month, are WhatsApp (94 percent) and Facebook (87 percent), followed by YouTube (76 percent), Instagram (67 percent), Facebook Messenger (66 percent), and Twitter (50 percent). Our analysis in this case study focuses specifically on the use of Facebook and WhatsApp. [198]

WhatsApp

WhatsApp is by far the most popular social media and communication platform used by Nigerians with smartphones. It is an inexpensive messaging service that takes little data to operate. In addition to its affordability, encryption, the ability to easily communicate with large groups, and send audio notes makes it user-friendly for populations who lack formal education or do not communicate in a written language.[199]

---
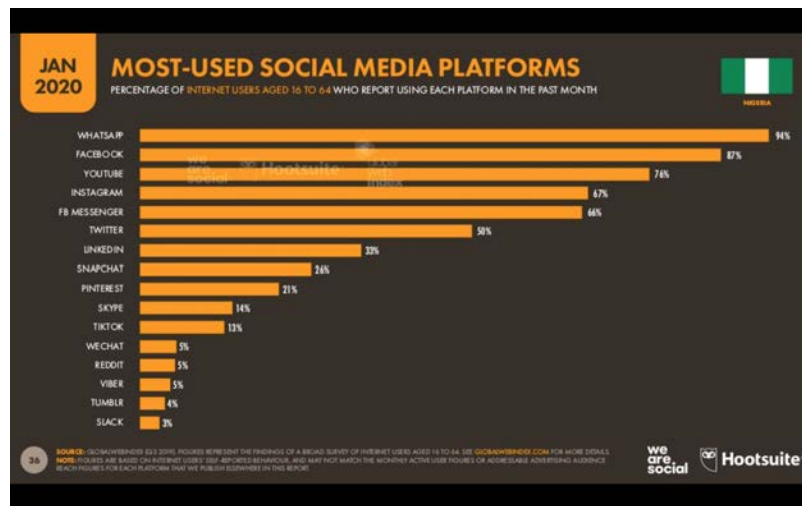
[198] Kemp, "Digital 2020," January 30, 2020.

[199] "How WhatsApp Is Used and Misused in Africa," The Economist, July 18, 2019, https://www.economist.com/middle-east-and-africa/2019/07/18/how-whatsapp-is-used-and-misused-in-africa.

The ability to share and forward information easily to other members within trusted groups is important to Nigerian social media users. Political, religious, and ethnic divisions as well as lack of trust in formal institutions cause Nigerians to place significant trust in the sender rather than the source of information. Therefore, it is more likely that information will be believed and shared if it is received from a known speaker or source within the recipient's social circle. Many Nigerians have extensive social networks and, as a result, belong to multiple large WhatsApp groups. Current rules limit the size of chat groups to a maximum of 256 people per group. [200]

WhatsApp's forwarding feature allows users to easily forward messages to multiple group chats at a time, allowing information to spread quickly through vast chains of social networks. Once a message moves through five message chains or is at least five forwards away from its original sender, it is labeled with a double arrow that indicates to the receiver they are viewing a message that has been mass-forwarded. Once a message is labeled with the double arrow, it can only be forwarded to one chat at a time to help "slow down the spread of rumors, viral messages, and fake news."[201]

Facebook

Facebook is more popular in Nigeria today than anywhere else in Africa, reaching more than 24 million users. Its inclusion in the Free Basics program was instrumental to Nigeria becoming Facebook's largest market on the continent. The ability to access Facebook for free makes it accessible to a broad audience, while platforms like YouTube and Instagram that require data and are known to have a more affluent and elite user base.[202]

Facebook was launched in Nigeria in 2016 when Mark Zuckerberg traveled to Lagos on his first trip to sub-Saharan Africa to unveil the new Free Basics program in partnership with Airtel Africa, the nation's second largest mobile phone network. As part of the promotion, Facebook was included in Airtel's free basics service package as one of the

---

[200] "What We've Learned about Fake News in Africa," BBC News, November 12, 2018, https://www.bbc.com/news/world-africa-46138284; "How WhatsApp Is Used and Misused in Africa."

[201] "WhatsApp FAQ - About Forwarding Limits," WhatsApp, accessed May 12, 2020, https://faq.whatsapp.com/en/android/26000253/.

[202] Yomi Kazeem, "More People Use Facebook in Nigeria than Anywhere Else in Africa," Quartz Africa, February 6, 2016, https://qz.com/africa/611516/more-people-use-facebook-in-nigeria-than-anywhere-else-in-africa/; Hassan and Hitchen, "Driving Division?," 6.

80 pre-selected websites that Airtel customers could access without having to pay for data.[203]

1.3 LAWS

Nigeria has adopted a robust legal framework that presumably protects free speech rights and prohibits hate speech (including many forms of dangerous speech) on traditional media and social media platforms. Constitutional provisions, international treaties, and national laws, including the 2010 Electoral Act and the 2013 Political Party Code of Conduct, explicitly prohibit the use and dissemination of speech intended to provoke sectional feelings or violent reactions.

In 2015, a new law was adopted in response to concerns over the nefarious use of social media and its increasing penetration. The 2015 Cybercrimes (Prohibition, Prevention, Etc.) Act criminalizes various harmful conduct within Nigeria's cyber space, including online hate speech on social media platforms. However, in practice, the 2015 Cybercrimes Act has not been applied to the most heinous perpetrators of online dangerous speech, but instead used to muzzle journalists and civil society organizations who attempt to report on government misbehavior. The Act was used by both local and state governments to arrest opponents and critics for alleged hate speech.[204]

At the time of writing this report, the Nigerian government is considering two additional bills that would further criminalize hate speech and misinformation shared on social media. The *National Commission for the Prohibition of Hate Speech* bill and the *Protection from Internet Falsehood and Manipulation and other Related Offences* bill would introduce additional penalties up to and including death for misuse of social media platforms.

The proposals have received harsh criticism from civil society groups and human rights organizations like Amnesty International, who claim the bills' vague language could be arbitrarily applied to censor critics. They note that many of the crimes described in the

---

[203] "What We've Learned about Fake News in Africa."

[204] "Cybercrimes (Prohibition, Prevention, Ect) Act, 2015" (2015), https://www.cert.gov.ng/ngcert/resources/CyberCrime__Prohibition_Prevention_etc__Act__2015.pdf; Peter Nkanga, "How Nigeria's Cybercrime Law Is Being Used to Try to Muzzle the Press," Committee to Protect Journalists, September 21, 2016, https://cpj.org/blog/2016/09/how-nigerias-cybercrime-law-is-being-used-to-try-t.php.

new bills are already covered by the 2015 Cybercrimes Act and fear additional punitive measures will further stifle Nigeria's media landscape. [205]

PART TWO: Social Media as a Weapon

Facebook and WhatsApp have influenced the media landscape in Nigeria for both better and worse. Per their design, these platforms are extremely effective communication tools, allowing users to stay in touch with far-reaching networks and share messages easily, quickly, anonymously, and cheaply. On one hand, these platforms allow Nigerians greater access to information, freedom of speech, and opportunities to improve government accountability, particularly in rural areas.[206] For example, the #BringOurGirlsHome media campaign brought international attention to the kidnapping of more than 200 schoolgirls by Boko Haram in 2014.

However, the same characteristics that contribute to social media being an effective advocacy tool also make it vulnerable to exploitation by various bad actors who promote dangerous online speech in the form of disinformation. Disinformation intentionally aims to manipulate or mislead the audience, while misinformation is often unintentionally spread on social media by unknowing audiences. Both have the capacity to intensify or exacerbate identity-based conflict.

Politically affiliated groups are increasingly using social media to spread mis- and disinformation in attempts to gain the upper hand during elections. These messages often play into long standing divisions in Nigerian society, especially between the Muslim north and the Christian south. Political parties in Nigeria are based on identity-based allegiances which lend themselves to a stark "us" versus "them" divide between different parties. These divisions are exacerbated by the zero-sum character of elections which determine who has most access and control over the country's resources. As a result, parties formed along ethnic lines commonly "exploit the benefits of ethnicity, religion, and similar patrimonial cleavages to gain political power."[207] Social media platforms are not responsible for the politicization of religion and

---

[205] "Nigeria Is Considering Incredibly Harsh Punishments for Social Media Users Who Criticize the Government," Amnesty International, December 4, 2019, https://www.amnesty.org/en/latest/news/2019/12/nigeria-bills-on-hate-speech-and-social-media-are-dangerous-attacks-on-freedom-of-expression/.

[206] Hassan and Hitchen, "Driving Division?"

[207] Christian Chukwuebuka Ezeibe and Okey Marcellus Ikeanyibe, "Ethnic Politics, Hate Speech, and Access to Political Power in Nigeria," *Africa Today* 63, no. 4 (2017): 70, https://doi.org/10.2979/africatoday.63.4.04.

ethnicity in Nigerian, but they have become an increasingly important venue for the most radical and dangerous elements to reach a broader audience.

The emergence of "content curators" who are paid by various political interests to create and spread disinformation has intensified political divisions and inflamed identity-based conflicts. Many of these paid partisans exploit existing division and mistrust within Nigerian society, as well as spread polarizing messages that are intended to manipulate a targeted group of voters. Content creators often create fake accounts on social media, distort images, and intentionally spread false information on social media.[208]

In the last five years, foreign firms paid for by Nigerian politicians and business leaders have spearheaded disruptive rumor campaigns to advance their own agendas. In 2015 a wealthy Nigerian supporter of then-President Goodluck Jonathan paid Cambridge Analytica to produce a media campaign to prop up the President prior to the elections that year. The firm, which was later involved with the UK's Brexit vote and the 2016 American presidential election, used this opportunity to pilot a strategy that capitalizes on divisive online messaging. It produced a campaign video that claimed that if Buhari should win the 2015 election he would impose Sharia law and sign a peace deal with Boko Haram, something anathema to the vast majority of Nigerians.[209]

The slickly produced video included a number of gruesome and disturbing images intended to intimidate and suppress northern voters while appealing to Christians in the south. While Buhari still won the election, the video was seen by millions of people. In addition to Cambridge Analytica, Israeli intelligence firms with links to Mossad, Israel's national intelligence agency, are also known to have had contact with Nigerian politicians participating in weaponization of social media.[210] Without regulation, Nigeria risks becoming a laboratory for further weaponization of social media.

---

[208] Jamie Hitchen et al., "How WhatsApp Influenced Nigeria's Recent Election — and What It Taught Us about 'Fake News,'" Washington Post, February 15, 2019, https://www.washingtonpost.com/news/monkey-cage/wp/2019/02/15/its-nigerias-first-whatsapp-election-heres-what-were-learning-about-how-fake-news-spreads/.

[209] Carole Cadwalladr, "Graphic Video Used by Cambridge Analytica to Influence Nigerian Election," The Guardian, April 4, 2018, https://www.theguardian.com/uk-news/2018/apr/04/cambridge-analytica-used-violent-video-to-try-to-influence-nigerian-election.

[210] Yomi Kazeem, "Facebook Shut down a Network of Fake Accounts That Targeted African Elections—from Israel," Quartz Africa, May 16, 2019, https://qz.com/africa/1621249/facebook-shuts-down-fake-news-accounts-targeting-africa-elections/.

As social media has been weaponized in Nigeria over the last five years, a number of tactics have emerged.

Misleading Headlines

Due to constraints on mobile data in Nigeria, many people who consume their news on social media only read the headlines and rarely read the actual article. Content creators and partisans have taken advantage of this by using inflammatory headlines that do not actually reflect the content of the article. For example, in 2017 the prominent online newspaper Vanguard ran an article with the headline "Mrs. Buhari feeds to death 18-month-old baby with kerosene." [211]

Considering that the President's last name is also Buhari, many people believed the article was referring to the President's wife. However, the article itself states that the mother, a kerosene seller in Benin-city named Mrs. Ata Buhari (no relation to the President or his wife), accidentally gave her child kerosene thinking it was water. This headline perpetuates a rumor that the President and his wife abuse children despite the fact that they were in no way involved in the tragic accident. This is an especially dangerous tactic because of Facebook's free app that allows users to read story headlines for free but does not provide access to the rest of the article.

Misleading Visuals

Content creators and political activists commonly use mislabeled images in their articles to spread their messages or narratives. This includes using misleading captions, such as falsely describing the actors in a photo, or using a photo from another context but claiming it illustrates violence described in the attached article. Sometimes this is the result of journalistic malpractice or mistakes. In other causes it is more manipulative. An example of how this can happen comes from an article accusing Boko Haram of using improvised explosive devices (IED) in a 2018 attack in Edo state.[212] The photo portrays a large explosion in rocky terrain However, as Nigerian civil society group CITAD pointed out, this photo is actually from the Islandic volcano explosion in

---

[211] Simon Ebegbulem, "Mrs Buhari Feeds to Death 18-Month- Old Baby with Kerosene - Vanguard News," Vanguard, February 19, 2017, https://www.vanguardngr.com/2017/02/mrs-buhari-feeds-death-18-month-old-baby-kerosene/.

[212] Editor, "IED Explosion from 'Suspected Boko Haram Members' Kills One in Edo," Daily Correspondents, January 17, 2018, http://dailycorrespondents.com.ng/2018/01/ied-explosion-from-suspected-boko-haram-members-kills-one-in-edo/.

2010. [213] In addition to eroding trust in journalists, in some cases, false, inflammatory images have caused further violence to occur

Mis-attributing Quotes

Another common tactic exploited by nefarious actors that is also a byproduct of poor journalistic standards is the misattribution of quotes. In Nigeria, this has been weaponized to discredit famous figures or shape a particular narrative around them. While many of the most egregious inaccurately-claimed comments from political and business leaders–including the President–are usually quickly proven false, many people still believe the initial false reporting if it fits their worldview.

For instance, in March 2020, it was reported that the former President Obasanjo denied coronavirus was present in Nigeria and had insulted the Health Minister. It later emerged that Obasanjo had never made these comments; in fact, they had been entirely fabricated. The lie was likely intended to investigate rumors that the current President Buhari's government was lying about the spread of the virus. Although they were disproven, the inaccurate reporting continued to spread on social media platforms.

Conspiracy Theories

Content creators and political activists have established a pattern of planting and amplifying conspiracy theories which build off truthful historical and current experiences, as well as the population's pre-existing beliefs about the country. For example, during the 2018 Presidential election the leader of the Indigenous People of Biafra, a designated terrorist by the Nigerian government, intentionally spread a rumor that President Buhrani had not left the country for medical treatment, but instead had been turned into a Sudanese clone.

Since most people in Nigeria have only experienced poor-quality healthcare services, many did not believe official reports that President Buhari had recovered from a period of illness and believe he must have been cloned. Additionally, another reason this rumor stuck was that Nigeria's former President Umaru Musa Yar'Adua had travelled to Saudi Arabia for health reasons before an election in 2009 and was not seen until his death six months later. Conspiracy theories not only spready patently false information

---

[213] Y.Z. Ya'u and Hamza Ibrahim, "Fake Images as Dangerous Speech: An Advisory" (Kano, Nigeria: Center for Information Technology and Development, May 2018), 8.

but exacerbate the public's tenuous trust in the government and became fodder for partisan propaganda.[214]
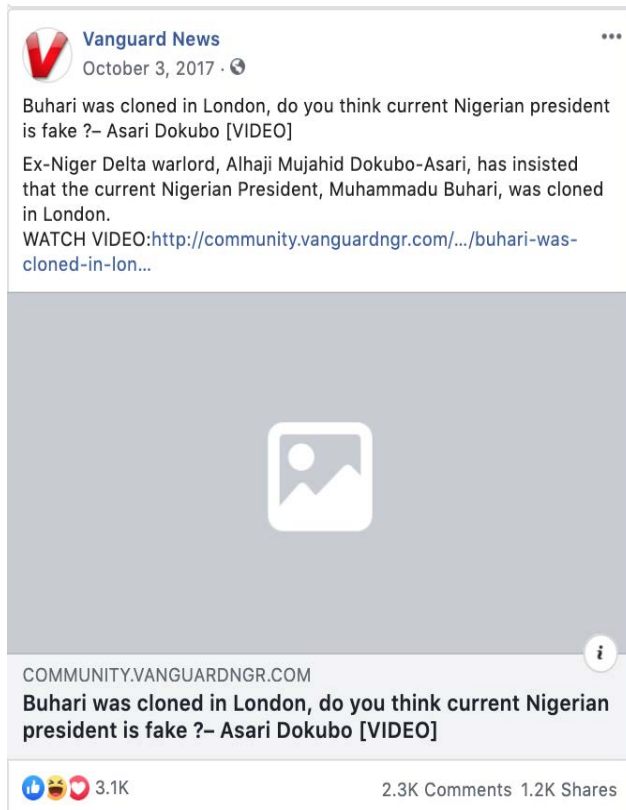
Cross-Posting

In some cases, content creators generate fake news sites or Facebook pages or impersonate established news agencies like *Vanguard Nigeria,* a prominent national newspaper, in order to spread disinformation. Some of these fake pages have more than 20,000 followers. These pages can become platforms for other types of dangerous speech tactics mentioned in this section.

2.2 DRIVERS OF DANGEROUS ONLINE SPEECH

A number of factors contribute to creating an environment in Nigeria where dangerous online speech can flourish.

---

[214] Hassan and Hitchen, "Driving Division?," 9.

Dangerous speech is incentivized by elites

A 2017 survey revealed 47 percent of respondents believe "political interest" was the main cause of dangerous speech.[215] Rather than being discredited, the use of dangerous speech and disinformation are integral components of political strategies to win elections in Nigeria. Having close ties to a politician is often the only way to gain financial stability in Nigeria, and political patronage has created a market for disinformation. Researchers found that "Nigerian political elites employ hate speech to curry favor along every possible line of diversity—such as ethnicity, geopolitical region, religion, and state of origin—to retain or obtain political power."[216] As a result, a cottage industry of "propaganda secretaries" has emerged, paid by elites to produce fake news that will benefit their particular party officials.[217]

Inherent trust in the "in-group" creates an enabling environment for disinformation

There are currently not enough trusted alternative sources of information available to factcheck disinformation spread on social media. Due to decades of censorship, ruling parties using the media as a mouthpiece of the state, and poor journalistic standards that allow false information to be re-reported as news, there are few options and little public trust in the Nigerian media. In fact, 22 percent of Nigerians surveyed in 2017 believe that journalistic malpractice was to blame for the emergence of fake news.[218]

In addition, family and social ties are incredibly important in Nigerian society.[219] Social connections are often lifelines, partially as a result of the state's failure to provide security and social services. Social media has enabled much of the population to maintain ties with members in their shared political, religious, or ethnic networks. However, lack of trust in the media and government officials means that information sent from a known member of a social circle is more likely to be believed than

---

[215] Terfa T. Alakali, Hemen Philip Faga, and Jinatu Mbursa, "Audience Perception of Hate Speech and Foul Language in the Social Media in Nigeria: Implications for Morality and Law," *Academicus International Scientific Journal*, no. 15 (2017): 169.

[216] Christian Chukwuebuka Ezeibe and Okey Marcellus Ikeanyibe, "Ethnic Politics, Hate Speech, and Access to Political Power in Nigeria," 66.

[217] "How WhatsApp Is Used and Misused in Africa."

[218] Alakali, Faga, and Mbursa, "Audience Perception of Hate Speech and Foul Language in the Social Media in Nigeria."

[219] Chika Okoye, "Social Networks in Urban Nigeria. Book Review of: Kate Meagher, Identity Economics. Social Networks and the Informal Economy in Nigeria," *Articulo - Journal of Urban Research*, 2013, http://journals.openedition.org/articulo/2360.

information received from outside it, even if there is little or no evidence supporting the claim.

High connectivity and low media literacy allow messages to spread quickly

Many Nigerians, young and old, remain unsure of how to identify "fake news" or report it. The overall literacy rate in Nigeria is 62 percent; 71 percent for men and only 53 percent for women. In 2019, the Economist quoted a Nigerian communications professor saying, "Most people who post have no idea what they're posting or reposting because they themselves are not media literate… They don't understand motives, they don't understand intent, they don't understand underflowing currents of other things."[220] WhatsApp, in particular, is commonly where older Nigerians unintentionally spread misinformation. A 2019 report from the Business and Human Rights Resource Centre "attributes that tendency to share as being down to a lack of digital literacy and a reliance on trusted social networks."[221]

With Nigerians belonging to large WhatsApp groups with the ability to forward messages to up to five chat groups at a time, information, whether true or false, spreads quickly through social networks. Many experts argue "the only resilient defense is savvier, more skeptical users."[222]

Laws against online hate/dangerous speech are not known or enforced.

Research from 2017, two years after the passage of the 2015 Cybercrime Act, shows that although Nigerian's believe that spreading hate speech and misinformation should have legal consequences, they are unaware of existing laws or punishments for doing so. Specifically, "All the participants agreed that there is need to take legal measures to regulate hate speech and foul language online in Nigeria; they however, denied having knowledge of any existing legal measure already in place to curb the menace of hate speech in Nigeria."[223]

---

[220] "How WhatsApp Is Used and Misused in Africa."

[221] Yomi Kazeem, "WhatsApp Is the Medium of Choice for Older Nigerians Spreading Fake News," Quartz Africa, October 22, 2019, https://qz.com/africa/1688521/whatsapp-increases-the-spread-of-fake-news-among-older-nigerians/.

[222] "How WhatsApp Is Used and Misused in Africa."

[223] Alakali, Faga, and Mbursa, "Audience Perception of Hate Speech and Foul Language in the Social Media in Nigeria."

The lack of public awareness about hate speech laws despite the overwhelming and ongoing presence of dangerous speech occurring in the media highlights the government's failure to properly and equitably apply the law. While elites and elected officials continue to use dangerous speech, the only people targeted with hate speech laws were political adversaries. The selective application of the law has contributed to a culture of impunity in regard to the spread of dangerous speech. In addition, the resulting lack of accountability further undermines public trust and reinforces "in-group" reliance, which deepens ethnic divides and intensifies communal conflicts.

2.3 RESPONSES OF KEY STAKEHOLDERS

Online dangerous speech is a known problem.

Findings from the 2017 survey on *Audience Perception of Hate Speech and Foul Language in the Social Media in Nigeria* reflect 85 percent of Nigerians agree or strongly agree that "hate speech and foul language is prevalent on social media." The report states that participants in the study suggest social media companies devise improved means for blocking and removing hateful speech, which is perpetrated by a minority at the expense of wider society. It was also reported that awareness must be raised about the need to maintain civility and tolerance.[224]

On multiple occasions, members of the Nigerian government have publicly addressed the need to curb mis- and disinformation. In 2017, Information Minister Lai Mohammed spoke repeatedly about the danger of fake news, warning that it could "threaten and destroy" the country. In 2018, the Nigerian government announced a nationwide campaign to raise awareness about the problem of fake news. Currently, the government is openly fighting back against health misinformation regarding the COVID-19 pandemic.[225]

Police and army personnel in Plateau State where communal conflict is ongoing are monitoring Facebook for fake information and fictitious pictures. A BBC report from 2018 states that in addition to monitoring and debunking false stories on Facebook, they hold regular meetings with local imams, pastors and politicians to raise awareness of the threat of misinformation in the region.

---

[224] Alakali, Faga, and Mbursa, 168–70.

[225] "Hold Purveyors of Fake News Accountable, Lai Mohammed Tells Nigerians," Vanguard, August 21, 2017, https://www.vanguardngr.com/2017/08/hold-purveyors-fake-news-accountable-lai-mohammed-tells-nigerians/; "Fake News and Nigeria's Herder Crisis," *BBC News*, June 29, 2018, sec. Africa, https://www.bbc.com/news/world-africa-44655148.

Facebook's response to calls to improve oversight of its platform in Nigeria has been limited.

Facebook currently has offices in 70 countries around the world, but only one on the continent of Africa.[226] The Africa office is located in Johannesburg, South Africa, and was opened in 2015.

Facebook currently partners with three local fact-checking organizations in Nigeria to review potentially false or dangerous content: Africa Check Nigeria, AFP Nigeria, and Dubawa.[227] However, these organizations do not appear to have the capacity to adequately review the high volume of speech shared on this platform. In 2018, BBC reported that two of Facebook's third-party fact-checking partners, AFP, and Africa Check, had committed just four full-time staff totals to analyzing and removing false news on the platform which is utilized by 24 million Nigerians. In addition, they found that Facebook's monitoring efforts in languages other than English are largely ineffective. A 2020 report by the Centre for Democracy and Development titled *Driving Division?* reported:

---

[226] "Our Mission," About Facebook, accessed May 12, 2020, https://about.fb.com/company-info/.

[227] "What Publishers Should Know About Third Party Fact Checking on Facebook," About Facebook, accessed May 12, 2020, https://www.facebook.com/business/help/182222309230722.

"Furthermore, platforms such as Twitter and Facebook do not have sufficient capacity to assess content, particularly when it is in local languages such as Igala or Kanuri. Even when they do have capacity in languages like Hausa, the number of people responsible for monitoring content is unable to keep up with the vibrant conversations (and abuse) taking place online."[228]

Facebook claims that it is addressing the issue of media literacy in Nigeria. In addition to their fact-checking initiative and machine learning tools, it has recently launched an "online safety and digital literacy youth programme" with 140 Nigerian secondary schools. However, Nigeria has more than 54,000 secondary schools as well as millions of Nigerian children who are not in school at all. More recently, in response to COVID-19, Africa Check launched a new initiative on WhatsApp called "Kweli," which means "truth" in Kiswahili.[229] Through this chatbot, anyone can submit information to be verified or fact-checked by WhatsApp for free.

Because dangerous speech and mis- and disinformation are driven and fueled in part by mistrust of political leaders and institutions, it is important that established, trusted, local organizations and community leaders are engaged in efforts to combat it.

PART THREE: How Dangerous Online Speech Exacerbates Communal Conflict between Farmers and Herders

Social media platforms in Nigeria are being used by political and business elites to further their agendas. This process is facilitated by low media literacy and lack of trust in formal institutions. Mis- and dis-information spread on social media play on pre-existing religious and ethnic tensions and are contributing to growing threats to Nigerian national security. Social media platforms are hosting and facilitating the spread of content that imbues relatively isolated local disputes with toxic religious and ethnic narratives of an existential threat thus providing a catalyst for escalatory violence.

---

[228] Hassan and Hitchen, "Driving Division?," 38.

[229] "How WhatsApp Is Used and Misused in Africa"; Halima Gikandi, "Nigeria Fights Fake News about Coronavirus," PRI The World, March 6, 2020, https://www.pri.org/stories/2020-03-06/nigeria-fights-fake-news-about-coronavirus.

Historically, herders and farmers have coexisted relatively peacefully.[230] However, as rainfall has become more erratic, herders have increasingly brought their livestock further south, competing with farming communities for access to land and water. At the same time, land tenure agreements have eroded and the mechanisms that previously handled conflicts undermined. A decade of fierce communal conflict in the 2000s, the growth of criminal cattle rustling networks, the proliferation of light arms, the rise of Boko Haram, and widespread impunity for perpetrators on all sides have normalized violence. In this context isolated alterations over land escalate and militias mobilize leading to clashes and retaliatory violence. Perpetrators on all sides largely enjoy impunity, and victims rarely receive any kind of justice.

Meanwhile, as news of the most recent attacks spreads, they become imbued with toxic narratives of Fulani genocidal jihad or a centuries-old effort by Christians to undermine Islam in Nigeria, linking horrific yet locally rooted violence to larger existential questions about the nation's identity. While dangerous online speech is rarely implicated in the initial violent incident, social media platforms play host to dangerous speech *after* an initial act of violence, deepening divisions and in some cases inspiring further conflict. This cycle has resulted in a series of escalatory tit-for-tat attacks which can claim hundreds of lives.

Using the dangerous speech paradigm to analyze online speech around conflicts between farmers and herders' sheds light on how this speech affects the evolving conflict. These narratives take complicated and local conflicts over land, indigeneity, and impunity and brings it all under the label "farmer-herder," which has come to signify a one-sided narrative of "Fulani herdsmen" who are solely perpetrators. This increasing demonization of Fulani plays into larger narratives of Islamic victimization in independent Nigeria, and risks pulling this part of Nigeria into what has happened in other parts of the Sahel where many Fulani have joined armed groups in response to perceived victimization at the hands of the state and other groups.

---

[230] Catherine Ver Eecke, "Pulaaku: Adamawa Fulbe Identity and Its Transformations" (Philadelphia, USA, University of Pennsylvania, 1988), http://search.proquest.com/docview/303719401/abstract/875E4BBB31764051PQ/1.

3.1 Underlying Drivers of Conflict

Hundreds of different ethno-linguistic communities who practice a mix of agriculture and pastoralism live in this area. The main pastoral group is the Fulani, who are one of the largest ethno-linguistic groups in West Africa.[231] While some Fulani still move seasonally with their livestock, most are sedentary and practice a blend of farming and pastoralism. While the Fulani are increasingly talked about as a monolithic group, there are actually multiple different Fulfulde language-groups within Nigeria and within these subgroups society is divided.[232] Across West Africa conflict between different elements of the Fulani community is just as, if not more, common than conflict between Fulani and other ethnic groups.

The vast majority of Fulani are Muslim and are often associated with the introduction of the religion to the region. In Northern Nigeria Usman dan Fodio led a revolution of Fulani and marginalized Hausa against pagan Hausa elites and established the Sokoto Caliphate in 1802. The Caliphate launched raids and expanded into what became central Nigeria until it was absorbed under the British colonial administration at the end of the 19the century. Many of the groups on the receiving ends of those raids, smaller farming communities such as Tiv, Berom and, Bachama, instead, converted to Christianity in the 19th and 20th centuries after contact with European missionaries. While none of these groups represent a heterogeneous ethnic unit—indeed conflict within groups is just as fierce and between groups—as violence has escalated, group boundaries have started to become more salient. This history has become fertile soil for the propagation of message which reflect mange of the hallmarks of dangerous speech.

After a decade of good rains after independence in 1960 the region has experienced waves of drought and desertification. Data from the early 2000s suggested the desert was moving south at a rate of 0.6 kilometers a year. According to the Nigerian Ministry of the Environment, between 50 percent and 75 percent of the northern-most states, with a population of 35 million people, are under severe threat of desertification.[233]

---

[231] There are between 25 - 40 million Fulani (also known as Fulbe, Puel, Fula) across West Africa stretched from Senegal to Sudan. Despite their large presence across the region they are divided by language (there are multiple Pulaar languages some of which may be mutually intelligible by native speakers) and colonial-era boundaries. In no country do they make up more than 40 percent of the population.

[232] Ver Eecke.

[233] "Great Green Wall for the Sahara and Sahel Initiative - National Strategic Action Plan" (Abuja, Nigeria: Federal Republic of Nigeria Ministry of Environment, October 2012), 17–18.

Pastoralists have adapted to the changing climate by migrating further south than they used to and by keeping their herds there for longer periods of time—sometimes even grazing their livestock there permanently. This has led to a sharp uptick in disputes over land and water rights with the growing population of sedentary agriculturalists.[234]

Land Tenure and Dispute Resolution

Over the last few decades the Nigerian Federal Government, with assistance from foreign development partners, has provided significant material aid to farming communities. These development projects have helped farmers expand their production to feed Nigeria's growing population.

At the same time, many of the over 400 grazing reserves established for herders in the 1960s have been appropriated by private commercial interests or encroached upon by growing urban settlements. Since 2017, a number of states including Taraba and Benue have instituted grazing bans, effectively making herding illegal and significantly escalating the tensions between herders and the state governments. The state-supported expansion of agriculture and undermining of pastoralism has led to heated competition for increasingly scarce resources, and a feeling among many pastoralists that they cannot trust the government.

Historically, disputes between pastoralists and agriculturalists were regulated by community appointed leaders while intractable issues would be sent to sharia courts, which were maintained in a hybridized form, until the end of the colonial era.[235] With independence the state began to take on more of these tasks, but this just resulted in police and judicial harassment of herders, who found the new systems inaccessible and foreign. After the return to democracy in 2000 political leaders, seeing more value in sedentary communities who can be relied on to show up at the polls, increasingly favored farming communities. As previous conflict resolutions mechanisms were undermined, those that filled their place were viewed as biased and illegitimate, increasingly leading people to seek redress outside the law.[236]

---

[234] "Herders Against Farmers: Nigerias Expanding Conflict" (International Crisis Group, September 2017), 3.

[235] Abdulmumini Adebayo Oba, "The Sharia Court of Appeal in Northern Nigeria: The Continuing Crises of Jurisdiction," *The American Journal of Comparative Law* 52, no. 4 (2004): 861, https://doi.org/10.2307/4144468.

[236] "Herders Against Farmers: Nigerias Expanding Conflict," 6.

Persistent Insecurity

Protracted insecurity has also contributed to conflict between farmers and herders. Between 2001 and 2010, more than 7,000 people were killed in communal conflict This pitted religious and ethnic communities against each other, led to massive population movements, and resulted in ethnic cleansing of large suburban areas. The killings increased the salience of ethnicity and religion in everyday life and deepened rivalries over land and access to political power by including deepening feuds between Christians and Muslims. While this violence was eventually forestalled in urban areas by the sustained deployment of the Nigerian military, it has continued sporadically in rural areas. Some scholars argue that conflicts between farmers and herders are a continuation of this violence including the crucial political dimensions.[237]

The Boko Haram insurgency in the north east has led tens of thousands of people to flee their homes, including many herders. The conflict has led to the closure of various grazing areas and livestock corridors, leading many pastoralists to move further south. As the militant group has sought to reform the politics of the Lake Chad basin through violence it has articulated an exclusionary, chauvinist, and violent narrative that draws on North Eastern Nigeria's unique political and religious history.[238] While there is little evidence that Fulani herders are working with Boko Haram, the presence of Boko Haram and their narrative and discourse have radicalized parts of Nigerian Muslim social media, while also leading to prominent Muslim leaders calling them out. At the same time, the horrific attacks perpetrated by the group are highlighted by partisans on the other side to allege that all Muslims, including President Buhari, are sympathetic to the militant group.

Over the last decade as the price of meat has risen, cattle rustling has become a lucrative criminal enterprise. According to a report from Nigerian civil society coalitions, in 2013 alone, more than 64,750 cattle were stolen and at least 2,991 herders killed in the north-central zone.[239]These criminal networks have taken advantage of the proliferation of small arms flowing from Libya. Some also blame the increase in cattle rustling on growing tensions within the Fulani community.[240]

---

[237] Jana Krause, personal communication.

[238] Alexander Thurston, *Boko Haram: The History of an African Jihadist Movement* (Princeton University Press, 2017), 3–11.

[239] Samuel Egwu, "The Political Economy of Rural Banditry in Contemporary Nigeria," in *Rural Banditry and Conflicts in Northern Nigeria*, n.d., 21–24.

[240] Aliyu Dahir Aliyu, April 25, 2020.

In response to the threats to their livelihood many herders have themselves purchased weapons and organized to protect themselves from cattle rustlers. This has been interpreted by some agriculturalist communities as a threat to their existence, who in response have also armed themselves.[241]

3.2 Dangerous Online Speech and conflicts in Adamawa and Plateau

A deeper investigation of two specific episodes of violence between so-called herders and farmers shows how dangerous speech on social media shapes violence in Nigeria. The two incidents were picked as larger incidents that attracted significant attention on social media, and also incidents in which people from both sides were victims. In both incidents' dangerous online speech heightened tensions after an initial altercation tying it into larger dangerous narratives that spawned deeper divides and, in some cases, led to further violence.

The first series of incidents under investigation took place in Numan local government area in northern Adamawa state in November and December of 2017. This area, which was on the periphery of the Sokoto Caliphate and also the site of significant missionary activity during the colonial era, is around 50 percent Christian and 50 percent Muslim. In 2003 and 2004, conflicts related to misunderstanding around a murder and the reconstruction of a mosque (burnt to the ground in 2003) led to clashes that left dozens dead and increased the salience of ethnicity and religion for many residents.[242]

Then, on November 20th, 2017, a group of Bachama attacked a handful of Fulani villages killing at least 80 people. Many of the dead were women and young children. Graphic images of the bodies were shared widely afterwards, and a week after the incident prominent associations and leaders from both sides began escalating rhetoric. A botched police operation in a Fulani village on December 1st led to widespread rumors of imminent retaliation which spurred hundreds of people to hurriedly flee Numan town. Three days later, eight Bachama villages were attacked by Fulani herders. The Nigerian air force intervened with a jet and attack helicopters. 86 people, including many civilians, were killed, and 3,000 houses were destroyed.[243]

---

[241] "Stopping Nigeria's Spiralling Farmer-Herder Violence" (International Crisis Group, n.d.), 4–6.

[242] Niels Kastfelt, "The Politics of History in Northern Nigeria," 2006, 166–79.

[243] "Harvest of Death: Three Years of Bloody Clashes Between Farmers and Herders in Nigeria" (Abuja, Nigeria: Amnesty International, December 17, 2018), 25–27; "Nigeria: Analysis of the Air Force Raid in Adamawa State" (Amnesty International, January 30, 2018); Published, "Suspected Herdsmen Ambush, Kill Six Police Officers in Adamawa," *Punch Newspapers* (blog), accessed April 13, 2020, https://punchng.com/suspected-herdsmen-ambush-kill-six-police-officers-in-adamawa/.

The second series of incidents under investigation took place in Plateau State in the local government area of Barkin Ladi and the city of Jos. Both areas were the sites of significant communal conflict in the early 2000s, and, in parts of Jos, people are still rebuilding after what amounted to ethnic cleansing.[244] Through May and June 2018, a series of small attacks by suspected herders led to heightened tension and increasingly dangerous statements from association leaders. Then between the 22 and 24th of June, over 200 people were killed in a series of coordinated attacks by suspected herders. During the attacks, misleading images spread to the city of Jos leading to mobilization which claimed the lives of 11. After the attacks, they were framed by many on social media as part of a larger religious conflict.[245]

Using Facebook to Spread Hate

The expansion of social media and the ability to spread graphic and emotive images and video has contributed to spirals of violence. In both Numan LGA and Plateau, state graphic images were taken with cell phones immediately after the violence and spread widely on social media, directly contributing in the case of Jos to the spreading of the violence. These disturbing images show people, including children and the elderly, with gruesome wounds from machetes, charred bodies, and mass graves. In some cases, the images that are spread are mislabeled – as in the images are sent saying "this is from x" but it is really from y. A cursory search of public Facebook posts reveals these images are shared and commented on widely and disproportionately to other posts in the aftermath of violence.[246]

As Fulani communities were burying their dead from the initial attacks on Fulani communities in Numan LGA in November 2017, dozens of photos were taken with cellphone cameras, uploaded to Facebook, and shared widely. These photos are incredibly graphic, showing lines of dead young children with horrific machete wounds.

On Facebook, most of these images were blurred out, but could be viewed by clicking a button in the bottom right. The spreading of these images was so prolific and seemed enough of a danger that it was even mentioned by the head of CAN in his

---

[244] Jana Krause, *Resilient Communities: Non-Violence and Civilian Agency in Communal War* (Cambridge University Press, 2018), 174–91.

[245] "Harvest of Death," 60.

[246] "Like. Share. Kill. Nigerian Police Say 'Fake News' on Facebook Is Killing People," BBC News, accessed April 10, 2020, https://www.bbc.co.uk/news/resources/idt-sh/nigeria_fake_news.

public statements after the attacks.[247] While these images were not directly implicated in future violence, their graphic nature caught the attention of a number of prominent individuals who then made inflammatory statements, adding to the spiral of conflict. They also have contributed to a sense of victimization by Fulani.

In Plateau, state graphic and misleading images were directly implicated in expanding violence from Barkin Ladi LGA to the suburbs of Jos in June of 2018. While violence was ongoing in Barkin Ladi and Riyom, images labelled as coming from Barkin Ladi of people with horrific machete wounds, mass graves, and murdered children were being spread on social media, especially in the city of Jos, 30 miles to the north. Some of the posts alleged that the conflict had moved to the outskirts of Jos and that the city was under imminent threat of attack.[248]In response groups of Berom men took to the streets and built roadblocks stopping all cars looking for Fulanis and Muslims. Eleven men were killed that day. However, a BBC Africa Eye investigation found that some of the images that were spreading were from previous incidents, or even other countries, and the reports that south Jos were under attack were false.[249] Many of these images are still online today.

Just as false statements made up quotes, and fake avatars have been used to advance political campaigns, there is evidence these same tactics have been used to deepen divisions following conflicts between farmers and herders. In the aftermath of the attacks on farming communities in Barkin Ladi in June 2018, *Premium Times* published an article in which they quoted a representative of Miyetti Allah saying:

"As much as I don't support the killing of human being, the truth must be told that those who carried out the attacks must be on revenge mission." The article was shared widely on social media, with many partisans calling this the smoking gun and saying Miyetti Allah was justifying killing people for cattle. However, a few days later the Premium Times retracted the story and said their journalist had misquoted the head of Miyetti Allah.[250] By this time, the message had already spread and was contributing to

---

[247] Kabiru R. Anwar, "How Numan Bloodshed Claimed 55 Women, Children," Daily Trust, November 26, 2017, https://www.dailytrust.com.ng/how-numan-bloodshed-claimed-55-women-children.html; "Villagers Attack Fulanis In Numan Adamawa, Kill 45, Injure Others (Graphic Pics)," Nairaland, November 21, 2017, https://www.nairaland.com/4189863/villagers-attack-fulanis-numan-adamawa.

[248] "Like. Share. Kill."

[249] "Like. Share. Kill."

[250] Samuel Ogundipe, "We Did Not Describe Plateau Killings as Retaliatory - Miyetti Allah," Premium Times, June 29, 2018, https://www.premiumtimesng.com/news/headlines/274275-we-did-not-describe-plateau-killings-as-retaliatory-miyetti-allah.html.

the narrative of Fulani aggression. According to Nigerian analysts, once information is spread, even if it is later proven to be false, it often stays online, and many people continue to believe it and its implicit messages.
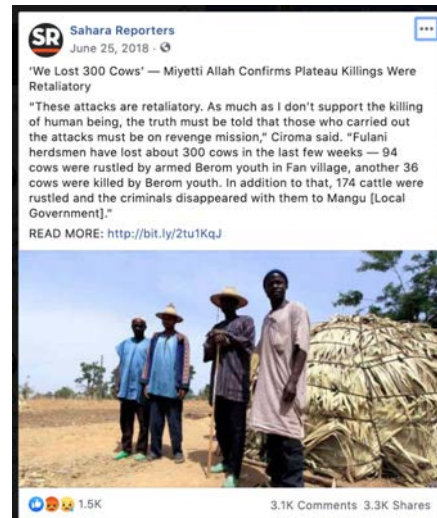
Tropes in Dangerous Online Speech

Dangerous online speech deepens divides and exacerbates conflicts between farmers and herders into an existential struggle through a series of tropes and motifs such as dehumanization, chosen trauma, justificatory mechanisms, accusations in the mirror, policing in-group behavior. These hallmarks of dangerous speech are found throughout social media during and after the events in Numan LGA in late 2017 and Plateau state in mid 2018.

Demonizing the Fulani

In both Numan LGA in 2017 and Plateau state in 2018, speech on social media that perpetuated a narrative of inherent Fulani savagery and a plot to take over the country. This narrative, which portrays a Fulani monolithic group of jihadists and terrorists and solely as aggressors, is rife in Nigerian media.[251] The label "killer herdsmen" has taken hold and is used in many online newspaper headlines whenever Fulani are involved,

---

[251] "Farmer-Herder Conflict: The Politics of Media Discourse in Nigeria," *Il Ponte*, January 2020.

even as victims.[252] This mischaracterized the conflict, obscuring the fact that there is significant tension within the Fulani community that results in violence among Fulani and that many Fulani are victims of cattle rustlers and ethnic militias themselves. It justifies any violence against Fulani as against an aggressive enemy while also obscuring the structural causes of the conflict.

After the initial attacks on Fulani communities in Numan in November 2017, a number of Nigerian newspapers used headlines and images that were misleading about the attacks. For instance, the Guardian said "At least 30 killed in herdsmen clash," which does not convey the fact that it was the "herdsmen" who were killed.[253] In the weeks and months following the initial attack on the Fulani settlements, and the subsequent attack on Bachama villages, a number of Nigerian newspapers used a photo of someone dressed in a boubou with a cap and a large machete chasing another man.[254] To a West African audience the man with the machete is clearly a Muslim and likely Fulani. The image paints the Fulani as ruthless and aggressive. The only problem is the photo is not from Numan. It is not even from Nigeria. The photo was taken years earlier in the Central African Republic, and as a reverse image search reveals, has been used in hundreds of articles about Fulani across West Africa.[255]

[252] Innocent Chiluwa Chiluwa Isioma M., "'Deadlier than Boko Haram': Representations of the Nigerian Herder–Farmer Conflict in the Local and Foreign Press - Innocent Chiluwa, Isioma M Chiluwa," *Media, War & Conflict*, February 7, 2020, http://journals.sagepub.com/doi/10.1177/1750635220902490.

[253] "At Least 30 Killed in Herdsmen Clash | The Guardian Nigeria News - Nigeria and World NewsNigeria — The Guardian Nigeria News – Nigeria and World News," accessed May 8, 2020, https://guardian.ng/news/at-least-30-killed-in-herdsmen-clash/.

[254] Hindi Livinus, "Herdsmen Attacks: We Feel like Foreigners, Second-Class Citizens in Our Own Land–Adamawa Victims," Spring, February 10, 2018, https://springng.com/herdsmen-attacks-feel-like-foreigners-second-class-citizens-land-adamawa-victims/.

[255] "Deadly Nigeria Conflict? No, Pic Was Taken in Central African Republic," Africa Check, accessed April 13, 2020, https://africacheck.org/spot-check/deadly-nigeria-conflict-no-pic-was-taken-in-central-african-republic/.

These misleading images and headlines simplify a complicated and nuanced situation with dangerous stereotypes that only reinforce preconceived notions that contribute to conflict.

In the aftermath of the attacks in Barkin Ladi and Jos, a number of partisans took to social media to allege the attacks were part of the alleged Fulani jihad that was backed by President Buhari. In a two-part Facebook post right after the attacks titled "UNDERSTAND THE FULANI JIHAD," one partisan called the recent attacks "genocide" before saying that President Buhari, is "a conscienceless JIHADIST who is here to execute a full-blown Jihad against the Non-Caliphate rest of Nigerians." He says the President, with the support of the Fulani in government, is giving herders an open reign to terrorize and "actualize the dream of dan Fodio," the founder of the Sokoto Caliphate.[256] While it is clear that the attacks in Barkin Ladi in June 2018 were perpetrated by groups of herders, there is no evidence it is part of a larger plan to take over southern Nigeria and forcibly convert people to Islam. This narrative takes local conflicts and makes them existential issues for people on the other side of the country who would otherwise not be particularly affected.

Justifying Violence

A range of justificatory mechanisms are mobilized to justify violence which serves to deepen divisions and sometimes inspire future violence. These tropes serve to condone violence, reducing in-group pressure for tolerance and understanding, and in some cases using previous incidents of violence to justify further attacks.

In the news coverage that spread on social media after the attacks on Fulani communities outside Numan town in November 2017 a number of the quotes in from farmers or Christians frame the attacks as reactions to previous Fulani transgressions. A youth leader in Numan told the Nigerian online news Daily Trust that the violence began when "a group of Fulani herders killed a Bachama farmer and injured his brother when they complained that their animals entered their farm." The district head brought up the contentious issue of indigeneity when he told reporters from the Daily Nigerian

---

[256] "Charles Ogbu - UNDERSTANDING THE FULANI JIHAD (PT1) By Charles...," Facebook, June 25, 2018, https://www.facebook.com/charles.ogbu2/posts/10156748163834255.

that there was an "unusual influx of the Fulani into our villages … It is compulsory for any settler to obtain a written permission from the traditional ruler of Numan chiefdom … none of these settlers can say he obtained permission."[257]
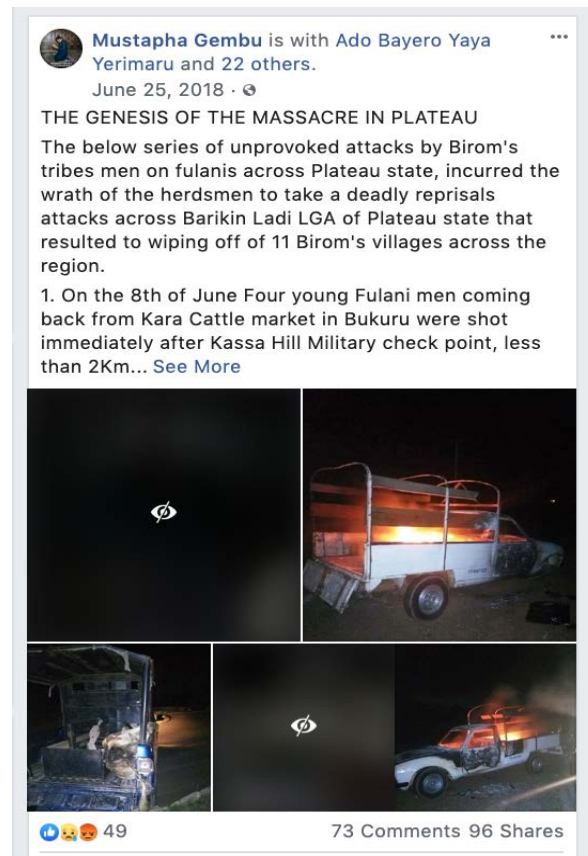
It is impossible to ascertain the veracity of these claims from afar, but the casual justification for the killings speaks to eroding norms against non-violence among citizens.

After Fulani herders killed over 200 people in Barkin Ladi over three days in June of 2018, a number of Muslim partisans took to social media to justify the violence as a response to alleged attacks of Fulani herders earlier in the month. Painting the victims of violence as guilty of some other crime is a common trope in ideological justification for violence.[258]

---

[257] Anwar, "How Numan Bloodshed Claimed 55 Women, Children"; "Bachama Militia Kill 45 Fulani Children, Women in Adamawa [GRAPHIC PHOTOS]," *Daily Nigerian* (blog), November 21, 2017, https://dailynigerian.com/bachama-militia-kill-45-fulani-children-women-in-adamawa-graphic-photos/.

[258] Jonathan Leader Maynard, "Rethinking the Role of Ideology in Mass Atrocities," *Terrorism and Political Violence* 26, no. 5 (2014): 14–15, https://doi.org/10.1080/09546553.2013.796934.

According to one Facebook post, which laid out six incidents of cattle rustling and/or assassination against Fulani in June alone, a "series of unprovoked attacks by Birom's tribes men on Fulanis across Plateau state, incurred the wrath of the herdsmen to take a deadly reprisals attacks across Barikin Ladi LGA of Plateau state that resulted to wiping off of 11 Birom's villages across the region." This line is picked up by partisans on Facebook, in some cases using the comments section of gruesome photos of dead Christians to write about the alleged crimes committed against Fulani. By justifying the attacks

Chosen Trauma

There is evidence some of this has evolved into a "chosen trauma," a term coined by social psychologist Vamik Volkan whereby a memory of group suffering, loss, helplessness, and humiliation in a conflict with a neighboring group becomes an important keystone moment in their identity, and in some cases a justification for violence.[259] Sharing images and reigniting the memory of previous victimization can serve as justification for future violence. In the months after the attacks on Fulani communities in Numan LGA, these images have helped turn these attacks into a "chosen trauma" for Fulani and Muslim partisans on social media. On the Fulani News Media on Facebook the images are brought up a few times after the incident as evidence of Fulani victimization. The images are also still the pinned item for the "Stop the Genocide Action Group", a Facebook page with 14,000 followers.[260] As incidents get cast in an incendiary light on social media, they have become fodder for future violence.

Accusations in the Mirror

There is evidence that partisans on both sides have propagated rumors of an impending attack in order to advance their own narrative or justify their own mobilization.

---

[259] Vamik D. Volkan, "Transgenerational Transmissions and Chosen Traumas: An Aspect of Large-Group Identity," *Group Analysis* 34, no. 1 (March 1, 2001): 79–97, https://doi.org/10.1177/05333160122077730.

[260] "Stop Genocide Action Group - HACKED TO DEATH WITH MACHETES AND KNIVES," Facebook, August 17, 2019, https://www.facebook.com/SGAGNGO/photos/a.485670971770154/965571547113425/?type=3&theater; "Fulani News Media - HACKED TO DEATH WITH MACHETES AND KNIVES; ARREST AND PROSECUTE THE KILLERS," Facebook, October 7, 2019, https://www.facebook.com/pg/fulaninewsmedia/posts/?ref=page_internal.

In Plateau state in June 2018, some of the posts alleged that the conflict had moved to the outskirts of Jos and that the city was under imminent threat of attack.[261] In response, groups of Berom men took to the streets and built roadblocks stopping all cars looking for Fulanis and Muslims. Eleven men were killed that day. The dynamic of rumors leading to mobilization is not necessarily new in Plateau state, as Jana Krause has shown how similar rumors and counter-mobilization were instrumental in mobilizing violence in Christian / Muslim communal conflict in Jos in the decade before Facebook came to Nigeria.

Questioning in-group Loyalty

Conflicts between farmers and herders have become a litmus test issue for policing in-groups and creating echo chambers on social media. An example comes from the comment section of a Facebook post of gruesome images of dead Christians in Plateau in June 2018. On June 24 (add year), an Igbo man posted a photo series of gruesome photos allegedly taken in Plateau days before. In the comment section, another Igbo man asks why the man posted the photos, considering that Igbo people are from the south-east and are not a party to the conflict in Barkin Ladi: "Why is it that it is Igbo people that are posting the pictures of Plateau killings? Are we involved in the killings?" And then later "something is wrong with Igbo. They are weeping more than the bereaved. People are playing politics with the lives of others and Igbo are fighting for them."

However, in response, a citizen-journalist (and a prolific Facebook user) who is also Igbo took screen shots of the comment and posted it, saying "I feel uniquely shamed, petrified and diminished to share the same ancestry" as the person posing the questions.[262] Essentially, the citizen-journalist, who contributes to many prominent online newspapers and has thousands of followers on Facebook, is implying that attacks on anyone who is Christian in Nigeria is an attack of Igbo people too and that

---

[261] "Like. Share. Kill. Nigerian Police Say 'Fake News' on Facebook Is Killing People," BBC News, accessed April 10, 2020, https://www.bbc.co.uk/news/resources/idt-sh/nigeria_fake_news.

[262] "Charles Ogbu - I Feel Uniquely Shamed, Petrified and Diminished To...," Facebook, June 25, 2018, https://www.facebook.com/charles.ogbu2/posts/10156748357334255.

should not be questioned. This silences internal dissent and paints it as traitorous to the in-group, increasing intolerance and radicalization while sealing off opportunities for dialogue and understanding.

Speaker

These messages have been spread by various leaders and communal associations, which have varying degrees of legitimacy, have nonetheless become seen as representatives of entire communities and increasingly become entangled in an unproductive cycle of sensational accusations and counter accusations which ratchet up tensions and perpetuate divisiveness.

As news of, and images from, the attacks on Fulani communities outside Numan town spread in November 2017, a number of Muslim representatives made various inflammatory statements heightening tensions between the two groups and providing fodder for social media propagandists to paint their enemies. On November 26, the Sultan of Sokoto, one of the cultural leaders in the north, while inaugurating the first Fulfulde radio station in Nigeria, warned people against misconstruing "patience for weakness," and said that what happened in Numan would not go unpunished.[263] The Jonde Jam Fulani Youth Association of Nigeria (JAFUYAN), a much smaller organization which claims to represent Fulani youths but in reality represent a vocal minority, released a statement where they said, "We also unequivocally demand justice for our killed members, as justice remains the only way to peace" and called the killings ethnic cleansing. In early December, the Chairman of the Muslim Council of Adamawa State denied that attacks were a result of farmer/herder conflict, but instead a calculated genocide against Muslims and Fulani.[264] While these utterances no doubt raised the tension among many Muslims, selective quotes were also spread by Christian partisans to as an example of alleged inherent Muslim aggression.

---

[263] Grace Alheri Abdu, "An Kaddamar Da Gidan Radiyon Fulfulde A Adamawa," Voice of America, November 27, 2017, https://www.voahausa.com/a/an-kaddamar-da-gidan-radiyon-fulfulde-a-adamawa/4137986.html; Iro Dan Fulani, "Adamawa Attack: Sultan Talks Tough; Warns against Killing of Fulanis," Premium Times, November 26, 2017, https://www.premiumtimesng.com/news/headlines/250648-adamawa-attack-sultan-talks-tough-warns-killing-fulanis.html.

[264] John Owen Nwachukwu, "Numan Killings: Deliberate Move to Eliminate Muslims, Fulani - Magaji," Daily Post Nigeria, December 3, 2017, https://dailypost.ng/2017/12/03/numan-killings-deliberate-move-eliminate-muslims-fulani-magaji/; Noah Ebije, "Numan Killings: Fulani Youths Spit Fire, Demand Justice," The Sun Nigeria, November 25, 2017, https://www.sunnewsonline.com/numan-killings-fulani-youths-spit-fire-demand-justice/.

In response Pene Da Bwatiye, a communal association claiming to represent the Bachama, issued a statement claiming there is no Bachama militia and that the use of "genocide" and "ethnic cleansing" were:

> "…weighty words/ phrases mischievously used to incite well-meaning Nigerians and tarnish the good name of our peace-loving people. Rather than keeping quiet therefore, we say without fear of contradiction that the world over, it is the Fulani herdsmen that established a terrorist militia which has been terrorizing farmers, eating crops and destroying livelihood all over Nigeria—believed to have killed more innocent people than their dreaded cousins in terror, the Boko Haram and yet without arrest."[265]

This message, spread as a press statement, was spread widely on social media. The message, coming for a group with a megaphone, even if they do not have many supporters, contains themes of dangerous speech and only exacerbated communal tensions in the area.

Countering Dangerous Speech

While social media is a conduit for dangerous online speech that has increasingly framed local conflicts in a national light and thus exacerbated the conflict, it is important to note that these platforms have also been used to spread counter speech. Days after the incident two members of the Miyetti Allah association told ATV Yola that everyone should wait to see what the government would do.[266] On Friday the 24th of November the chairman of CAN in Adamawa state condemned the perpetrators "in strong terms" and described it as a "wicked act," and emphasized the Christian commandment "thou shall not kill. He also referenced the gruesome photos that were being shared. By not endorsing the violence or even seeking to offer justification this is an example of counter speech.

---

[265] "Press Release by Pene Da Bwatiye on the Recent Herdsmen and Farmers Clash in Numan Local Government Area of Adamawa State 1st December 2017," TGNews, December 2, 2017, http://www.tgnews.com.ng/2017/12/02/press-release-by-pene-da-bwatiye-on-the-recent-herdsmen-and-farmers-clash-in-numan-local-government-area-of-adamawa-state-1st-december-2017/.
[266] "NUMAN FULANI," YouTube, November 21, 2017, https://www.youtube.com/watch?v=I6zXA80AnX8.

In Barkin Ladi, a story of rescue behavior by an Imam, who saved 262 Christians, was widely shared in the aftermath of the violence and later picked up by the BBC. As the attack in Barkin Ladi started people started fleeing to the house of the 83-year-old Imam Abdullahi Abubakar who hid the women at his house and locked the men in the mosque. When the attackers arrived and demanded he turn the people over, he prostrated himself and asked that they leave everyone in peace. The attackers complied. News of his courageous behavior spread widely on social media, and after a journalist from the BBC corroborated the story and did their own piece, Abubakar was recognized as a hero and given various awards in the hope that he can serve as an example to others. Social media platforms were used to spread an inspiring story of rescue behavior and humanity to spread widely amidst all the violence.

PART FOUR: Conclusion

Nigeria is a center of innovation and leadership on many critical issues taking shape across the continent. Despite only 14 percent of the population being active users, the weaponization of social media over the last five years has had a corrosive effect on the country, contributed to deepening divisions, and exacerbating communal conflict. This weaponization has been facilitated by inadequate public media literacy and a lack of resources and incentives for quality journalism, which has been simultaneously undermined by the advent of social media and repressed by the state. The inability of the state to provide basic services or security in some region's further breeds mistrust of formal institutions. While civil society groups have identified both the root and contributing causes of violence, social media platforms, the Nigerian government, and donors have yet to adequately embrace their proposed solutions. If left unaddressed, the future remains one where conflicts created by local factors – namely access to increasingly scarce natural resources – will remain vectors for dangerous online speech. In this bleak picture, it can be expected that social media platforms will continue to be used to seed religious and ethnic divisions, broaden participation in local disagreements, and inspire participation in future violence.

There is looming concern that conflicts between farmers and herders in Nigeria may become enmeshed with larger extremist movements in the Sahel.[267] Since 2012, grievances within Fulani communities in Mali have morphed from isolated attacks against local Fulani elites to affiliation with regional jihadist groups and large-scale attacks on the state. In response to a perceived threat, farming communities have

---

[267] Modibo Chaly Cissé, "Understanding Fulani Perspectives on the Sahel Crisis – Africa Center," Understanding Fulani Perspectives on the Sahel Crisis, April 22, 2020, https://africacenter.org/spotlight/understanding-fulani-perspectives-sahel-crisis/.

organized their own militias for protection, and have been implicated in preemptive attacks categorized later as "ethnic cleansing."[268] Governments have increasingly ascribed collective guilt to all Fulani people, which has become a recruiting tool for regional jihadist groups.

 While the structural conditions that have given rise to communal conflict involving Fulani herders are similar across the Sahel, the risk that they unite and pose a legitimate threat to the secular states of the region – even with the corrosive influence of social media - is low. Fulani people are divided by a multitude of languages, castes, and socio-economic status and vast majority are deeply invested in the current order. However, as the livelihood of an important segment of their society is undermined, and they are collectively ascribed guilt for the actions of a minority, individuals might seek help from other places in the Sahel.

News of massacres of Fulani people in Mali and Burkina Faso have been shared widely on social media across the Sahel, fueling a growing sense of marginalization and victimization. While the prospect of a regional conflict involving Fulani is extremely unlikely, the danger of disparate elements united by similar grievances organizing and recruiting in online echo chambers grows. Understanding the root causes of both Fulani's and their neighbors' grievances is imperative to stopping the cycle of violence and further radicalization. While addressing these issues is the most effective way to foster long-term peace, social media companies can do more to ensure their platforms are free from dangerous speech used to incite identity-based violence.

---

[268] Jean-Hervé Jezequel, "Centre Du Mali : Enrayer Le Nettoyage Ethnique," Crisis Group, March 25, 2019, https://www.crisisgroup.org/fr/africa/sahel/mali/centre-du-mali-enrayer-le-nettoyage-ethnique.

RECOMMENDATIONS

The response of governments and social media companies to the problem of online dangerous speech has been inconsistent at best and negligent at worst. Social media platforms are in a unique and complex position. On one hand, they provide a space for democratic principles of engagement, freedom of expression, and instill an equalizing potential for learning and opportunity. However, as data show, these products simultaneously advance dangerous ideologies and amplify the targeting of vulnerable populations, both for abuse and mobilization.[269]

Many states lack the legal infrastructure or political will to independently regulate online speech. As seen in Nigeria, Brazil, and the Philippines, social media usage has rapidly become interwoven into public policy discourse and political agendas. Social media has been successfully manipulated by public officials to leverage select campaigns and advance polarizing narratives. Further, it is the obligation of companies to regulate their product without defaulting solely to the norms and laws of a country partner.

There are not consistent and uniform regulatory structures aimed to ensure the efficacy and ethical usage of such products. The inconsistent enforcement of online speech laws mainly is due to the ambiguity of what constitutes dangerous speech and how it differs from personal expression, which can and should be protected.

However, as the case studies have illustrated, dangerous speech has serious societal consequences and increases the likelihood of further violence taking place, especially in societies without strong protections for persecuted minority groups, functioning legal mechanisms, and histories of conflict and exploitation. Issues of accountability for online speech have been raised by the International Criminal Court and Court of Justice[270], civil society organizations such as the Global Network Initiative, and many state governments around the world who are negatively affected by dangerous speech online. These groups have asked technology and social media companies (who profit from operating in developing countries with extremely high rates of social media

---

[269] T Camber Warren, "Explosive Connections? Mass Media, Social Media, and the Geography of Collective Violence in African States," *Journal of Peace Research* 52, no. 3 (2015): 297–311; Séraphin Alava, Divina Frau-Meigs, and Ghayda Hassan, "Youth and Violent Extremism on Social Media: Mapping the Research" (Paris, France: United Nations Educational, Scientific and Cultural Organization, 2017), https://unesdoc.unesco.org/ark:/48223/pf0000260382.

[270] Philippines, Myanmar and Bangladesh have had open investigations under the ICC with Facebook use identified as a key tactic in quickly spreading dangerous speech and misinformation, mobilizing violence offline. The investigations in the Philippines led to their departure from the Rome Statute in 2017, rendering the ICC unable to continue its investigations on the human rights violations committed during the country's war on drugs.

penetration) to do more to protect the safety of some of the most vulnerable populations around the world.

Facebook, a leader in this space, has made some efforts to curb dangerous speech and misinformation. For example, both Facebook and Twitter have increased efforts to stem the spread of misinformation regarding COVID-19; their proactive efforts have led to a new wave of content regulation where posts from public and political officials have been removed for spreading dangerous, misleading or misinformation.

Additional efforts by Facebook include:

- Enhancing Community Standards and pledging to "remove content, disable accounts, and work with law enforcement when we believe there is a genuine risk of physical harm or direct threats to public safety"
- Changing the Newsfeed algorithm to slow the spread of fake news through deprioritizing certain posts
- Pledging to take more punitive approach for administrators of toxic groups, and will factor in moderator behavior when assessing the health of a group
- Hiring third party fact checkers, internationally certified through reputable journalism schools. Fact checkers are local members of civil society or journalists who alert Facebook to harmful content, potential fake accounts, and, mis and disinformation campaigns
- Increasing digital literacy, targeting states with poor internet infrastructure. The program "We think digital" is a localized user training to increase safe use of the product[271]

At the same time, Facebook, Twitter, YouTube, and WhatsApp, among others, have consistently failed to uniformly limit dangerous speech and most mis- and disinformation online.[272]

Facebook acknowledges the problem posed by the high volume of dangerous speech spread on its platform, stating the company could "never hire or partner with enough human beings to monitor all of the news published on its platform." However, in many

---

[271] "Community Standards," About Facebook, 2020, https://www.facebook.com/communitystandards/.
[272] Emily Dreyfuss and Issie Lapowsky, "Facebook Is Changing News Feed (Again) to Stop Fake News," Wired, April 10, 2019, https://www.wired.com/story/facebook-click-gap-news-feed-changes/.

instances even minimal monitoring efforts have not been applied.[273] In particular, Facebook has shied away from punishing political and public figures for sharing dangerous speech on their platform. Political leaders and their supporters commonly use these platforms to spread false and inflammatory rhetoric which inflame societal divisions that contribute to outbreaks of communal conflict.
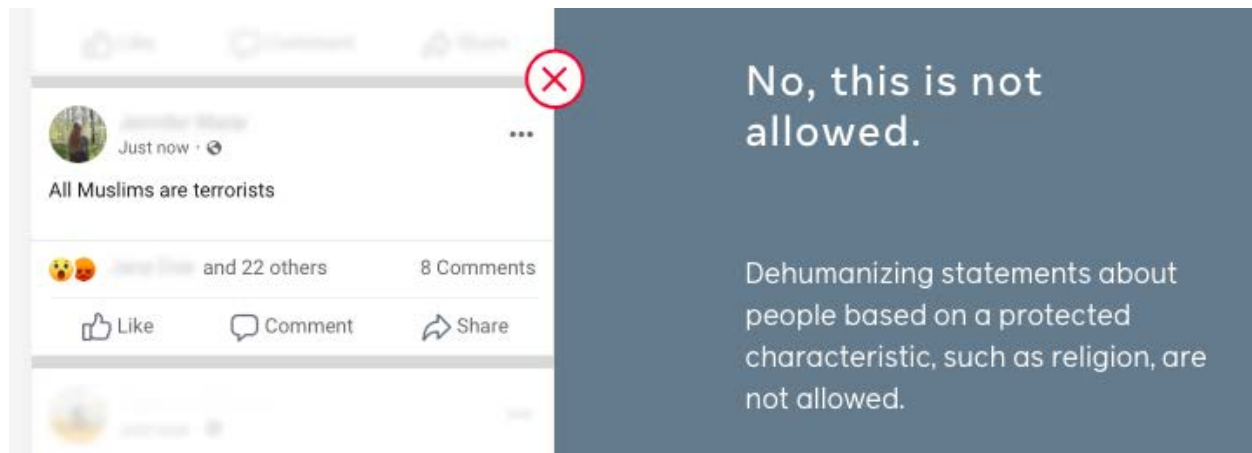
Facebook says dehumanizing speech is removed from the platform, but as proved in the enclosed case studies, this is not always the case.

Unlike public facing platforms, private mediums face additional regulatory obstacles. WhatsApp uses end-to-end encryption, ensuring consumer privacies unique to its messaging format. The medium has been connected to the spread of dangerous speech and misinformation through its forwarding feature, allowing the spread of otherwise private conversations. WhatsApp has moved to drastically reduce the volume of forward abilities, where it was once over 250 forwards per message, it is now five for a single, original message[274], and established message tracking mechanisms to inform consumers of a forwarded message's reach.

But the issue of transparency remains, WhatsApp does not know the content of

---

[273] Dreyfuss and Lapowsky; Katie Rogers et al., "Trump's Suggestion That Disinfectants Could Be Used to Treat Coronavirus Prompts Aggressive Pushback," The New York Times, April 24, 2020, https://www.nytimes.com/2020/04/24/us/politics/trump-inject-disinfectant-bleach-coronavirus.html.
[274] "WhatsApp FAQ - About Forwarding Limits," WhatsApp, accessed May 12, 2020, https://faq.whatsapp.com/en/android/26000253/. WhatsApp policy states, "when a message is forwarded through a chain of five or more chats, meaning it's at least five forwards away from its original sender, the message is labeled with a double arrow icon . These messages can only be forwarded to one chat at a time, as a way to help keep conversations on WhatsApp intimate and personal. This also helps slow down the spread of rumors, viral messages, and fake news."

messages in its platform, permitting illegal and inflammatory material to spread under its radar. To disrupt user misuse, WhatsApp has introduced digital literacy programs globally. Program implementation varies by state, and their enforcement of consumer education remains inconsistent.

Combatting the lack of uniformity in policy implementation, especially where state-sponsored actions are concerned, is a logical next step for companies. This trend of internet reliance is projected to increase as the world continues to modernize and globalize. Paired with the complexity of dangerous speech online, social media companies must invest in social media literacy. Yet, social media literacy is not enough.

Relying only on consumers to use social media appropriately is unrealistic and evades their responsibilities as a global corporation impacting billions of lives. Companies must, therefore, be aggressive in confronting dangerous social and political patterns of speech that erode trust and safety online, as when Twitter deleted tweets made by Brazilian President Bolsonaro touting false claims regarding the novel COVID-19 virus.[275]

Social media companies promote themselves as platforms that "build community." In order to preserve their reputation as platforms that bring the world closer together rather than forums of fake news and dangerous speech, they must invest in a way forward that addresses the negative impact of their products.

From the lessons learned in this report's case studies, the following policy recommendations have emerged with overwhelming urgency:

1. **Expand media literacy efforts.**
   1.1. Available data show that when users are armed with the knowledge and experience of recognizing disinformation and troll or bot communication patterns, they are more likely to reject any engagement or information therefrom.[276] Mandating literacy programs for consumers is an imperative commitment for companies striving to combat harm amplified on their product. Social media companies should partner with trusted grass-roots

---

[275] Raphael Tsavkko Garcia, "Harmful Tweets From High Places: Why Is Twitter Acting Now?," Undark Magazine, April 18, 2020, https://undark.org/2020/04/18/twitter-deleting-covid-19-lies-bolsonaro/.

[276] Se-Hoon Jeong, Hyunyi Cho, and Yoori Hwang, "Media Literacy Interventions: A Meta-Analytic Review," *The Journal of Communication* 62, no. 3 (June 1, 2012): 454–72, https://doi.org/10.1111/j.1460-2466.2012.01643.x.

organizations to promote media literacy and enhance counterspeech efforts.

    1.2.   Social media literacy programs should be culturally relevant and designed for key demographics' accessibility and comprehension.

    1.3.   Companies should go outside of their platform to increase social media literacy and build public sector and civil society partnerships to systematically integrate critical thinking and usage skills across a population. Sites for partnership may include schools, youth groups, and older adult housing facilities.

2.   **Enhance knowledge of country-specific contexts in which they operate to inform violence-prevention efforts.** Led by local offices with resident employees, companies should conduct the following in all sites of operation:

    1.1.   Utilize evidence-based research that has demonstrated regional or national periods of vulnerability, increased polarization and simultaneous surges in harmful content, and physical violence.

    2.1.   Implement country-specific "surge" reports that track when dangerous speech is most likely to be spread to enhance monitoring and counterspeech efforts.

    2.2.   Require language proficiency of content regulating staff for all widely spoken languages.

    2.3.   Content regulators and third-party fact-checking partnerships must be established before efforts are made to increase product penetration.

3.   **Implement and enforce global, uniform standards of online speech explicit to political leaders and public figures.**

    3.1.   Political advertisements should be limited and heavily regulated rather than expanded, with mandated, in country fact-checking prior to public viewing.

    3.2.   Improve accountability and transparency in algorithm regulation, diversity, and maintenance.

        3.2.1.   For example, companies should break down quarterly reports in real time and make them public and be fully transparent with post removals.

    3.3.   Apply unique regulations per the reach and usage by people in positions of power, namely political figures and their supporters and, ensure uniform enforcement of punitive measures through improved transparency (i.e. remove posts or shut down accounts of figures who spread dangerous speech).

   3.3.1. Assign content moderators and managers to track political candidate discourse, whether from a candidate directly or supporters.

  3.4. Globally, social media training led by respective companies for political candidates should be discontinued and discouraged.

4. **Participate in multilateral efforts and investments to forge the path forward.** Social media companies should look to the UN for human rights centered regulatory guidance and vice versa. There is a clear need to implement policies within the global landscape since much of the internet can be reached internationally. Technology company leaders cannot rely on states to regulate mediums undefined by their borders, although usership and trends are context specific, with culturally distinct norms and interpretations. In this respect, we recommend the following:

  4.1. A new body of multilateral institution cyber security task force on dangerous speech is needed to inform corporate social responsibility. The discourse on dangerous speech online is currently not uniform. The task force may be integral to set forth precedent and vision of norms countries can adopt into law and companies can apply to their platforms.

  4.2. This should result in a stronger base for global discourse on the uniformity of these standards, promote human rights as digital rights, and create consistent, context-specific, annual reports on fundamental dynamics of social media, speech, and state affairs.

5. **Reject calls to encrypt conversations on Facebook Messenger, which remains a major conduit of nefarious images and dangerous speech.**

  5.1. Should utilize technological advancements to screen messages for content that goes against community guidelines.

6. **Distinguish "Dangerous Speech" from "Hate Speech" in their community standards and regulatory practices.**

  6.1. Company policies pertaining to speech vary by platform. But as evidenced in this report, dangerous speech has been connected to violence and is by definition, different than hate speech. Doing so should inform and improve their policies on speech and increase action against dangerous speech.

These policy approaches address persistent gaps in the global regulatory landscape of social media at the nexus of dangerous speech. At present, the world has marginal

checks and balances in enforcement, consistency, and united standards online. Simultaneously, state-sponsored dangerous speech and disinformation are proliferating around the world.

As companies rush to create policies in an effort to reinforce scientific, evidence-based facts, and eliminate misinformation, and dangerous speech propelled on their platforms, the demand for more robust, integrated partnerships that value human rights are both essential and obligatory for the private sector to enact. In this unprecedented era, states have demonstrated unreliability in protecting freedom of expression while prioritizing the safety and wellness of those within their borders. It is, therefore, an international obligation to establish a uniform, multifaceted model informed by hyper-localization efforts in order to quell harmful dynamics online, especially those, which transfer to violence offline.