2022

The Anti-Human Rights Machine: Digital Authoritarianism and The Global Assault on Human Rights

Richard Ashby Wilson
University of Connecticut School of Law, richard.wilson@uconn.edu

The Anti-Human Rights Machine: Digital Authoritarianism and The Global Assault on Human Rights

Richard Ashby Wilson*

ABSTRACT

Across the world, governments and state-aligned actors increasingly target human rights defenders online using techniques such as surveillance, censorship, harassment, and incitement, which together have been termed "digital authoritarianism." We currently know little about the concrete effects on human rights defenders of digital authoritarianism as researchers have focused primarily on hate speech targeting religious, national, and ethnic minority groups. This article analyzes the effects of digital authoritarianism in two countries with among the highest rates of killings of human rights defenders in the world; Colombia and Guatemala. Anti-human rights speech in these countries portrays defenders as Marxist terrorists who are anti-patriotic and corrupt criminals. Evidence for a direct causal link to offline violence and killing is limited, however, and this empirical study documents the non-lethal and conditioning effects of speech. Human rights defenders who are targeted online report negative psychological and health outcomes and identify a nexus between online harassment and the criminalization of human rights work. Many take protective measures, engage in self-censorship, abandon human rights work, and leave the country. To prevent these harms, social media companies must implement stronger human rights-protective measures in at-risk countries, including expediting urgent requests for physical protection, adopting context-specific content moderation policies, and publicly documenting state abuses. The article concludes by advocating for a new United Nations-sponsored Digital Code of Conduct that would require states to adopt transparent digital policies, refrain from inciting attacks, and cease illegally surveilling human rights defenders.

I. INTRODUCTION

The online harassment and threats against Ramón Cadena Rámila began in 2018 when a vitriolic column in the Guatemalan daily newspaper *El Periódico* spread quickly on social media. Ramón Cadena is one of Central America's most prominent human rights attorneys and he has served as a judge on the Inter-American Court of Human Rights and represented indigenous and environmental activists opposing hydroelectric and mining projects. The Foundation Against Terrorism, a group representing military veterans of Guatemala's counterinsurgency war, coordinated a campaign on Twitter and Facebook labeling Cadena a Marxist, a liar, and a millionaire who grew rich by stealing his clients' monetary reparations won in litigation. Strangers accosted him and his family in restaurants, calling them thieves and threatening them with violence. Cadena requested government protection. Receiving none, he sent his teenage daughter to live abroad with family. He petitioned the Inter-American Commission of Human Rights which ordered the Guatemalan authorities to provide 24-hour armed police accompaniment. He suffers from anxiety, insomnia, acid reflux, and every day he takes a different route to his office at the International Commission of Jurists in Guatemala City. His concerns are well-founded: from 2018 to 2020, thirty human rights defenders were killed in Guatemala.[1]

Across the world, governments and state-aligned actors orchestrate online harassment campaigns against human rights defenders.[2] Government online propaganda operations have been termed "digital authoritarianism,"[3] and are characterized by an array of anti-democratic techniques that include internet shutdowns, surveillance,[4] censorship of online speech,

---

[1] FRONT LINE DEFENDERS, GLOBAL ANALYSIS 2019, at 4 (2020), https://www.frontlinedefenders.org/en/resource-publication/global-analysis-2019.

[2] "State actors" denotes government officials or agencies and "state-aligned actors" denotes individuals whose speech on social media aligns closely with the interests of the government or military. "Human rights defender" is defined by the United Nations Office of the High Commissioner for Human Rights here: https://www.ohchr.org/EN/Issues/SRHRDefenders/Pages/Defender.aspx.

[3] *See* ADRIAN SHAHBAZ, THE RISE OF DIGITAL AUTHORITARIANISM (2018) (discussing the rise of digital authoritarianism); EROL YAYBOKE & SAM BRANNEN, PROMOTE AND BUILD: A STRATEGIC APPROACH TO DIGITAL AUTHORITARIANISM (2020) (discussing policy responses to digital authoritarianism). Ron Deibert defines authoritarianism as "state constraints on legitimate democratic political participation, rule by emotion and fear, repression of civil society, and the concentration of executive power in the hands of an unaccountable elite." Ron Deibert, *Cyberspace Under Siege*, 26 J. DEMOCR. 64 (2015).

[4] David Kaye (*Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*), U.N. Doc. A/HRC/41/35 ¶26 (May 28, 2019).

disinformation, state-sponsored trolling,[5] and incitement of violence.[6] Governments coordinate online propaganda operations to intimidate and silence critics and to galvanize popular support for a range of restrictive measures that include the criminalization of human rights work, and in some contexts, disappearances and killings.[7]

This article is the first to provide a theoretical framing of state-aligned propaganda campaigns against human rights defenders based on quantitative and qualitative social science research methods. It combines analyses of the content of the speech directed at defenders with evidence on the impacts of digital authoritarianism on the professional work and personal lives of defenders in two countries where human rights defenders are at risk. Colombia and Guatemala have among the highest numbers of lethal attacks on human rights defenders in the world and in 2020 ranked number one and four respectively in the Americas,[8] and number one and seven respectively in the world.[9] In the two-year period from 2018 to 2020, 106 defenders were killed in Colombia and fifteen were killed in Guatemala.[10] Attacks on defenders increased during the COVID-19 pandemic and in 2020, there were 177 killings in Colombia and fifteen in Guatemala.[11] Visible physical harms are only part of the account, however, and detentions and threats also increased in both countries.[12] To comprehend the full picture of a hostile environment for human rights, researchers need to complement quantitative measures with qualitative research on the emotional and psychological harms that defenders experience. [13]

---

[5] "State-sponsored trolling" is the coordination by an official state agency or party in government of automated accounts (bots), paid users, and volunteers to harass opponents. *See* CARLY NYST & NICK MONACO, INST. FOR THE FUTURE, STATE-SPONSORED TROLLING: HOW GOVERNMENTS ARE DEPLOYING DISINFORMATION AS PART OF BROADER DIGITAL HARASSMENT CAMPAIGNS, 1 (2018) (defining and analyzing state-sponsored trolling).

[6] *See* MUNA ABBAS ET. AL, INVISIBLE THREATS: MITIGATING THE RISK OF VIOLENCE FROM ONLINE HATE SPEECH AGAINST HUMAN RIGHTS DEFENDERS IN GUATEMALA (2019) (describing digital authoritarianism in Guatemala); Tamar Megiddo, *Online Activism, Digital Domination and the Rule of Trolls*, 58 COLUM. J. TRANSNAT'L L. 394 (2020) (discussing digital domination of civil society organizations by governments).

[7] ABBAS ET. AL, *supra* note 6; Megiddo, *supra* note 6.

[8] *See* FRONT LINE DEFENDERS, GLOBAL ANALYSIS 2020, (2021), https://www.frontlinedefenders.org/en/resource-publication/global-analysis-2020.

[9] *Id.*

[10] FRONT LINE DEFENDERS 2019, *supra* note 1.

[11] FRONT LINE DEFENDERS 2020, *supra* note 8. At the height of the COVID-19 pandemic in March 2020, the populist president of El Salvador Nayib Bukele tweeted that "organizations of 'human rights'. . . are on the side of the virus." @nayibbukele, TWITTER (Mar. 29, 2020, 5:08 PM), https://twitter.com/nayibbukele/status/1244370925815988226?lang=en.

[12] *See* U.N. 75 Sess., 8749th mtg. U.N Doc SC/14252 1 (July 14, 2020).

[13] *See* Allison J. Pugh, *What Good are Interviews in Thinking About Culture?*

The article documents the online content and character of coordinated online campaigns against human rights defenders and provides a coding guide listing twelve categories of anti-human rights speech. Digital authoritarianism has similar characteristics globally that involve accusations that defenders are subversives or terrorists who are guilty of corruption and criminality, and it also varies according to the culture, history, and language of a country.[14] In the qualitative interviews conducted for this study, defenders report many damaging effects not currently captured in the official datasets, including fear and intimidation, reputational damage, negative health effects, the need to take protective security measures, and the suppression of their public speech.[15] Online attacks undermine human rights work on a daily basis, and in extreme instances defenders have fled their homes and applied for asylum in a foreign country.[16] A majority of defenders interviewed for this study identified a causal nexus between anti-human rights speech[17] online and real-world violence. The minority who refrained from linking speech to violence still emphasized the ways in which anti-human rights speech conditions a population to tolerate violence against defenders.[18]

These empirical findings anchor a theoretical framing that integrates the two dominant analytical models of online hate speech: what I term the "Minority Model" and the "Political Model." The Minority Model, currently the dominant theory of online hate speech in social science, seeks evidence for causation or correlation between online speech and physical attacks on immigrants and minority (religious, racial, ethnic, etc.) social groups.[19] Studies in the Minority Model have identified statistical correlations between

---

Demystifying Interpretive Analysis, 1 AM. J. CULT. SOCIOLOGY 42 (2013) (discussing the advantages of qualitative research); Richard Ashby Wilson, *The Digital Ethnography of Law: Studying Online Hate Speech Online and Offline.* 3 J. LEG. ANTHRO. 1 (2019) (noting the ethnographic study of social media hate speech).

[14] *See infra* Part IV.

[15] *See infra* Part V.

[16] *See infra* Part VI.

[17] Anti-human rights speech is defined fully in Part V, and includes, *inter alia*, threats, accusations of criminality and corruption, dehumanizing language, denigrating statements about gender or sexual orientation, and other forms of disparaging speech targeting human rights defenders or organizations.

[18] *See infra* Part VI.

[19] Social science studies in the Minority Model. *See* David Yanagizawa-Drott, *Propaganda and Conflict: Evidence from the Rwandan Genocide*, 129 Q.J. ECON. 1947, 1989 (2014); Scott Straus, *What Is the Relationship between Hate Radio and Violence? Rethinking Rwanda's "Radio Machete*," 35 POLITICS & SOCIETY 609 (2007); Karsten Müller & Carlo Schwarz, Fanning the Flames of Hate: Social Media and Hate Crime (Jun. 8, 2020) (unpublished manuscript); *see also* Griffin Edwards & Stephen Rushin, *The Effect of President Trump's Election on Hate Crimes* 6-7 (Jan. 18, 2018) (unpublished manuscript) (analyzing the effects of political speech during the US presidential elections).

online speech and offline violence, thus laying to rest the question of whether there are offline harms in online hate speech. However, they are limited by the assumption that social media is comprised of autonomous individual actors and have paid less attention to the networked, state-sponsored, and automated nature of social media hate campaigns. This framework needs to be combined with the Political Model of online speech that is salient in the law and policy literature and highlights how states and state-aligned actors commandeer social media to stigmatize and undermine alternative and dissenting voices.[20] In turn, studies in the Political Model could benefit from the hallmark of Minority Model research, namely, an empirical social science component that systematically documents the harmful consequences of online speech.

The theoretical framework guiding this study integrates both models to analyze online campaigns against human rights defenders who are challenging impunity for conflict-era crimes or combatting government corruption. Both approaches have their advantages; the Minority Model is attentive to questions of causation and direct incitement against religious, racial, and ethnic groups, and the Political Model addresses both physical harms and long-term societal effects. Whereas the Minority Model highlights visible and often spectacular acts of physical violence, the Political Model highlights non-lethal impacts, including fear, intimidation, and the disruption and silencing of human rights defenders.

In the two countries studied, the principal consequence of digital authoritarianism is not direct incitement of violence, although that does occur. Instead, digital authoritarianism is a core element of a government-aligned propaganda campaign to control the public narrative on past and present human rights violations, and to demoralize and silence civil society actors. It fosters an atmosphere of tolerance for coercive acts such as the criminalization of human rights work. Thus, the Political Model is the most appropriate for understanding coordinated attacks on defenders, but it needs to draw theoretical and methodological inspiration from the empirically oriented Minority Model.

The article concludes with a set of recommendations for social media platforms and national governments that draws from international human rights law. Social media companies must implement stronger human rights-protective measures in at-risk countries by creating more channels for urgent action requests for protection by defenders, adopting content moderation policies that are context specific, dismantling state-sponsored propaganda

[20] For policy studies in the Political Model *see*, *e.g.*, ABBAS ET AL., *supra* note 6; Megiddo, *supra* note 6; NYST & MONACO, *supra* note 5, and JONATHAN CORPUS ONG, JEREMY TINTIANGKO & ROSSINE FALLORINA , HUMAN RIGHTS SURVIVAL MODE: REBUILDING TRUST AND SUPPORTING DIGITAL WORKERS IN THE PHILIPPINES (2021).

networks, creating mechanisms to document state abuses, and moving away from a one-size-fits-all model of content moderation. The United Nations should develop a new Digital Code of Conduct that requires states to adopt transparency in their digital policies, to refrain from inciting attacks on individuals or groups, and to cease their illegal surveillance of human rights defenders.

II. THE RISE OF DIGITAL AUTHORITARIANISM

At first, many observers applauded the democratizing potential of social media. In 2011 and 2012, pro-democracy movements in Egypt, Syria, Tunisia, and Russia organized mass protests against authoritarian regimes on Facebook.[21] In Latin America, civil society activists quickly mobilized on social media against government corruption and human rights violations.[22] However, governments soon took to social media and adopted the same mass mobilization practices and a decade later, digital technologies often serve to consolidate state power.[23] Analysts have coined a variety of terms to describe the range of current online tactics pursued by governments, including; "digital repression,"[24] "digital domination,"[25] and "digital neo-colonialism."[26] Government propaganda is, of course, nothing new, but the immediacy and scale of large-scale surveillance on social media have fundamentally altered its character, complexity, and capacity.[27] That governments on every continent engage in covert propaganda campaigns online is no longer in doubt, and social media companies openly acknowledge this reality. For instance, Twitter regularly updates its "Information Operations" archive documenting widespread platform manipulation by

---

[21] *See* Deibert, *supra* note 3, at 65; Zeynep Tufekci, *Social Movements and Governments in the Digital Age: Evaluating a Complex Landscape*, 68 J. INT'L AFF. 1 (2014); Zeynep Tufekci, *How Social Media Took Us From Tahrir Square to Donald Trump*, MIT TECH. REV. (Aug. 14, 2018), https://www.technologyreview.com/2018/08/14/240325/how-social-media-took-us-from-tahrir-square-to-donald-trump/.

[22] LEOPOLDO FERGUSSON & CARLOS MOLINA, CEDE, FACEBOOK CAUSES PROTESTS (2019).

[23] *See* Tufecki, *How Social Media Took Us From Tahrir Square to Donald Trump*, *supra* note 23; Deibert, *supra* note 3, at 65; Sam Gregory, *Cameras Everywhere Revisited: How Digital Technologies and Social Media Aid and Inhibit Human Rights Documentation and Advocacy*, 11 J. HUM. RTS. PRACT. 373, 373–92 (2019).

[24] NYST & MONACO, *supra* note 5, at 9.

[25] Megiddo, *supra* note 6.

[26] William Gravett, *Digital Neo-Colonialism: The Chinese Model of Internet Sovereignty in Africa*, 20 AFR. HUM. RTS. L. J. 125 (2020).

[27] *See* JEN WEEDON, WILLIAM NULAND, & ALEX STAMOS, FACEBOOK, INFORMATION OPERATIONS AND FACEBOOK (2017) (providing Facebook's analysis of government information operations).

governments.[28]

Many of the forms of state surveillance, censorship, and political manipulation of social media that are prevalent today were first practiced by the Chinese Communist Party which created the archetypal model of digital authoritarianism.[29] This began as long ago as the late 1990s, when China launched its Golden Shield Project; a surveillance system integrating population databases, identification tracking systems, street surveillance cameras, and facial recognition software, to which it added digital surveillance tools.[30] The (in)famous "Great Firewall of China" blocks foreign content, censors speech, and restricts access to certain sites or the internet altogether.[31] These techniques were quickly adopted by countries in the Middle East and elsewhere.[32] Additionally, the Chinese government imprisoned social media users for violating vague rules against spreading "online rumors,"[33] and countries such as Turkey adopted similar repressive tactics against journalists and activists.[34]

From 2014 onwards, authoritarian regimes such as China and Russia shifted their tactics from restricting access and censoring content to coopting social media and flooding public discourse with pro-government propaganda. [35] The massive surplus of pro-government online speech was wielded as a "censorial weapon"[36] in the information wars, undercutting the ability of civil society organizations to engage in counter-speech and challenge dominant narratives. Initially, governments set up networks of automated accounts (bots) to amplify their message and create the appearance of popular grassroots support (also known as "astroturfing"), a notorious practice of

---

[28] Information Operations, TWITTER, https://transparency.twitter.com/en/reports/information-operations.html.

[29] Xiao Qiang, *President XI's Surveillance State*, 30 J. DEMOCR. 53 (2019); NYST & MONACO, *supra* note 5, at 8; Megiddo, *supra* note 6, at 10.

[30] Xu Xu, *To Repress or To Co-opt? Authoritarian Control in the Age of Digital Surveillance.* 65 *AM. J. POL. SCI.* 309, 310 (2021).

[31] *See* Qiang, *supra* note 29; Peter L. Lorentzen, *China's Strategic Censorship,* 58 *AM. J. POL. SCI.* 402 (2014); *Timeline: China and Net Censorship*, BBC NEWS (Mar. 23, 2010), http://news.bbc.co.uk/1/hi/8460129.stm; JACK GOLDSMITH & TIM WU, WHO CONTROLS THE INTERNET? ILLUSIONS OF A BORDERLESS WORLD (2006).

[32] Helmi Noman & Jillian C. York, *West Censoring East: The Use of Western Technologies by Middle East Censors 2010-2011*, THE OPENNET INITIATIVE 1 (Mar. 2011), http://opennet.net/west-censoring-east-the-use-westerntechnologies-middle-east-censors-2010-2011.

[33] Ben Blanchard, Hui Li & Paul Carsten, *China Threatens Tough Punishment for Online Rumor Spreading*, REUTERS (Sept. 9, 2013), https://news.yahoo.com/china-threatens-tough-punishment-online-rumor-spreading-100229793.html.

[34] NYST & MONACO, *supra* note 5, at 35.

[35] Deibert, *supra* note 3, at 65.

[36] Tim Wu, *Is the First Amendment Obsolete?* KNIGHT FIRST AMENDMENT INSTITUTE (Sept. 1, 2017), https://knightcolumbia.org/content/tim-wu-first-amendment-obsolete.

Russia's Internet Research Agency.[37] As platforms became more aggressive in removing bots, government information operations established pro-government youth groups to engage in "patriotic trolling."[38] Digital militias such as China's "50 Cent Army" flood social media with nationalist propaganda, disinformation, and angry rhetoric directed at their political opponents.[39] Smear campaigns against human rights defenders became common.[40] Digital militias do not only drown out opposition voices; by decentralizing a propaganda campaign, they also obscure the role of the state and permit "plausible deniability" by political leaders.[41] The tactics of digital authoritarianism are constantly transforming and have progressed from restricting information to manufacturing an overabundance of speech. Digital authoritarianism has moved from the firewall to the firehose, and from suppression to cooptation.[42]

As China and Russia transformed the central features of digital authoritarianism, many democratic government security services practiced intrusive surveillance, including famously by the National Security Agency (NSA) in the United States.[43] Surveillance of independent journalists and human rights activists is widespread and has become increasingly sophisticated with the advent of military-grade surveillance software

---

[37] *See* NYST & MONACO, *supra* note 5, at 31 (analyzing Ecuador's contracts with private companies to set up fake accounts); Megiddo, *supra* note 6, at 15; SAMANTHA BRADSHAW & PHILIP N. HOWARD, THE GLOBAL DISINFORMATION ORDER: 2019 GLOBAL INVENTORY OF ORGANISED SOCIAL MEDIA MANIPULATION 18 (2019); Marco T. Bastos, & Johan Farkas, *Donald Trump is my President! The Internet Research Agency Propaganda Machine*, 5 SOC. MEDIA + SOC'Y 1 (2019).

[38] BRADSHAW & HOWARD, *supra* note 37, at 9; NYST & MONACO, *supra* note 5, at 11; Tufecki, *How Social Media Took Us From Tahrir Square to Donald Trump, supra* note 23, at 7; Bulut Ergin & Erdem Yörük, *Digital Populism: Trolls and Political Polarization of Twitter in Turkey*, 11 INT'L J. COMM. 4093 (2017). Anne Henochowicz, *Youth Volunteers to Spread Sunshine Online*, CHINA DIGITAL TIMES (Apr. 13, 2015), https://chinadigitaltimes.net/2015/04/translation-youth-volunteers-to-spread-sunshine-online/; Arzu Geybulla, *In the crosshairs of Azerbaijan's patriotic trolls*, OPENDEMOCRACY (Nov. 22, 2016), https://www.opendemocracy.net/en/odr/azerbaijan-patriotic-trolls/.

[39] Gary King, Jennifer Pan & Margaret E. Roberts, *How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, not Engaged Argument*, 111 *AM. J. POL. SCI 484 (2017).*

[40] #SmearCampaign, FRONT LINE DEFENDERS GLOBAL, https://www.frontlinedefenders.org/en/violation/smear-campaign.

[41] Deibert *supra* note 3, at 69.

[42] Christopher Paul & Miriam Matthews, *The Russian "Firehose of Falsehood" Propaganda Model: Why it Might Work and Options to Counter It*, RAND CORP. (2016), https://www.rand.org/pubs/perspectives/PE198.html.

[43] Jack M. Balkin, *The Constitution in the National Surveillance State*, 93 MINN. L. REV. 1, 3–4 (2008) (describing the prevalence of state surveillance); Deibert *supra* note 3, at 75 (discussing NSA surveillance).

programs such as Pegasus that are currently available only to governments.[44] Surveillance is not without consequences for those being surveilled, and UN officials have drawn a causal connection between government surveillance and the detention and torture of activists, and "possibly" extrajudicial killings as well.[45]

The techniques of digital authoritarianism have spread to democracies in tandem with the rise of right-wing populism.[46] The number of populist governments worldwide has doubled since the advent of social media, and many populist leaders mobilized the constituencies through using graphic speech online, replete with crude insults, misogyny, racial resentment, and xenophobia.[47] Some of the practices of digital authoritarianism are also present in classic democracies, including South Korea,[48] the United Kingdom,[49] and the United States.[50] The line between the digital practices of democratic and authoritarian governments has become blurred, and researchers from the Computational Propaganda Research Project found evidence of "organized social media manipulation" by the government or a political party in eighty-one countries in 2020.[51] There is compelling

---

[44] *See* Kaye, *supra* note 4, ¶9 (outlining the use of Pegasus by forty-five governments to monitor individuals); Washington Post Staff, *Takeaways From the Pegasus Project*, WASH. POST (Aug. 2, 2021).

[45] David Kaye (*Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*), U.N. Doc. A/74/486 ¶ 1(Oct. 9, 2019).

[46] *See* RALPH SCHROEDER, SOCIAL THEORY AFTER THE INTERNET: MEDIA, TECHNOLOGY, AND GLOBALIZATION 60 (2018) (asserting that digital media were a necessary precondition for the rise of right-wing and nationalist movements in China, India, Sweden, and the US).

[47] *See* IVAN KRASTEV & STEPHEN HOLMES, THE LIGHT THAT FAILED: A RECKONING 20-23 (2019) (describing the recent rise of populist demagogues and authoritarianism); Ronald F. Inglehart, & Pippa Norris, *Trump, Brexit, and the Rise of Populism: Economic Have-Nots and Cultural Backlash* 3, Harvard Kennedy Sch. Paper RWP16-026, (2016); Jonathan T. Rothwell & Pablo Diego-Rosell, Explaining Nationalist Political Views (Dec. 29, 2017) (unpublished manuscript).

[48] BRADSHAW & HOWARD, *supra* note 37, at 1, 18; Barton Gellman & Laura Poitras, *U.S., British Intelligence Mining Data from Nine U.S. Internet Companies in Broad Secret Program*, WASH. POST (June 7, 2013).

[49] Glenn Greenwald & Andrew Fishman, *Controversial GCHQ Unit Engaged in Domestic Law Enforcement, Online Propaganda, Psychology Research*, THE INTERCEPT (June 22, 2015), https://theintercept.com/2015/06/22/controversial-gchq-unit-domestic-law-enforcement-propaganda/.

[50] Lloyd Grove, *How Breitbart Unleashes Hate Mobs to Threaten, Dox, and Troll Trump Critics*, THE DAILY BEAST (Apr. 13, 2017), https://www.thedailybeast.com/how-breitbart-unleashes-hate-mobs-to-threaten-dox-and-troll-trump-critics (noting how White House strategist Steve Bannon's use of partisan news outlets and social media to attack opponents); *see also* BRADSHAW & HOWARD, *supra* note 37, at 9 (describing a USAID-sponsored social media program in Cuba).

[51] SAMANTHA BRADSHAW, HANNAH BAILEY, & PHILIP N. HOWARD, UNIV. OXFORD,

evidence that digital authoritarianism intensified over the course of the COVID-19 pandemic as some governments used the public health crisis as a pretext to intensify surveillance and stifle freedom of expression online.[52] We can safely conclude that digital authoritarianism is now a generalized feature of nation-state governance.

Digital authoritarianism has been well-documented in the Political Model literature produced by policy research centers and international nongovernmental organizations that have provided detailed reports on state-sponsored harassment, surveillance, and censorship of journalists, civil society organizations, and human rights defenders. However, thus far there has been a dearth of social scientific research on the concrete effects of state-sponsored harassment and the existential consequences for human rights defenders. The theoretical frame adopted here aims to bridge the Political and Minority Models by documenting and analyzing the impacts on human rights defenders of state-sponsored information operations.

III. RETHINKING THE HARM IN HATE SPEECH

In the aftermath of the Rwandan genocide, political scientists and economists applied advanced econometric techniques to determine whether there was a causal connection between inciting broadcasts on Rwandan radio and mass atrocities.[53] With the rise of social media, social scientists adapted these methods to examine the relationship between online hate speech and hate crimes against religious, racial, and ethnic groups in North America and Europe. Studies in the Minority Model present evidence of a correlation between online hate speech and offline violence against minority groups. For instance, by analyzing over 500,000 posts and comments on the Facebook page of the political party Alternative for Germany, Karsten Müller and Carlo Schwarz find a statistically significant correlation between anti-immigrant and anti-Muslim posts and offline attacks on Muslims and immigrants in Germany in 2016.[54] Anti-refugee hate crimes were more prevalent in areas with higher exposure to anti-refugee sentiment online, and this was especially

---

INDUSTRIALIZED DISINFORMATION: 2020 GLOBAL INVENTORY OF ORGANIZED SOCIAL MEDIA MANIPULATION 1 (2020), https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/127/2021/02/CyberTroop-Report20-Draft9.pdf.

[52] Andy Wang, *Authoritarianism in the Time of COVID*, HARV. INT'L REV. (HIR) (May 23, 2020), https://hir.harvard.edu/covid-authoritarianism/; Kristine Eck & Sophia Hatz, *State Surveillance and the COVID-19 Crisis*, 19 J. HUM. RTS 603 (2020); Adrian Shahbaz & Allie Funk, *Freedom on the Net 2020: The Pandemic's Digital Shadow*, FREEDOM HOUSE (2020), https://freedomhouse.org/report/freedom-net/2020/pandemics-digital-shadow.

[53] Yanagizawa-Drott, *supra* note 19; Straus, *supra* note 19.

[54] Müller & Schwarz, *supra* note 19.

true for violent incidents such as arson and assault.[55]

Researchers have identified similar effects in the United States. In a study of 100 US cities between 2011 and 2016, Kunal Relia et al. find that hate crimes correlate with tweets containing targeted discrimination on the basis of race, ethnicity, and national origin.[56] Griffin Edwards and Stephen Rushin found that the inflammatory online rhetoric used by then-candidate Donald J. Trump in the 2016 presidential election was associated with a statistically significant increase in reported hate crimes, with the sharpest increases in counties that voted for Trump by the widest margins.[57] Free-speech advocate and Twitter CEO Jack Dorsey defended the social media company's decision in January 2021 to de-platform Trump as thus, "Offline harm as a result of online speech is demonstrably real."[58]

Having established a clear correlation between online hate speech and physical hate crimes, researchers then turned to isolating the mechanisms that could explain the relationship between the two. Müller and Schwarz found that Facebook's algorithm elevated hate posts on users' feeds, convincing them that anti-immigrant sentiment was much more widespread in Germany than it actually was.[59] Facebook's internal studies revealed that their algorithm, by elevating the visibility of posts that garner more attention, promotes violent speech and disinformation on users' feeds.[60] Taken together, these findings highlight the relationship between the prevalence of online hate speech and the business model of social media companies, based as it is on an "attention economy" that thrives on provocation, sensationalism, and outrage.

Important as these findings are regarding attacks on protected groups (racial, religious, ethnic, sexual orientation, disability, etc.), there is little research on attacks orchestrated by state or state-aligned actors on occupational groups such as journalists or human rights defenders. Attacking human rights defenders is a hallmark of populist and authoritarian regimes, but as yet social scientists have not investigated how governments are taking

---

[55] *Id.* at 19.

[56] Kunnal Relia et al., *Race, Ethnicity and National Origin-based Discrimination in Social Media and Hate Crimes Across 100 U.S. Cities,* ICWSM (2019) (testing the relationship between social media and hate crimes in the United States).

[57] *See* Edwards & Rushin, *supra* note 20, at 19.

[58] Adela Suliman, *Trump Asks Court to Force Twitter to Reinstate His Account*, THE WASH. POST (Oct. 2, 2021).

[59] Müller & Schwarz, *supra* note 19; *see also* Tufekci, *Social Movements and Governments in the Digital Age, supra* note 21, at 6.

[60] Keach Hagey & Jeff Horwitz, *Facebook Tried to Make Its Platform a Healthier Place. It Got Angrier Instead*, WALL ST. J. (Sept. 15, 2021), https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215; s*ee also* Force v. Facebook, 304 F.Supp. 3d 315 (2018) (Katzmann, J., dissenting) (describing how Facebook's algorithm amplified the inciting messages of Hamas).

advantage of the affordances of social media to undermine human rights work. Furthermore, existing studies of online hate speech mainly focus on Western Europe and North America and there is little research on the Global South and in languages other than English.[61] Current studies focus primarily on liberal democracies rather than societies with elevated levels of political violence, shaky rule of law, and a recent history of armed conflict. Next, studies in the Minority Model assume that individual users act independently of one another, and they do not account for the ways in which attacks are orchestrated by state and state-aligned agencies. State-backed propaganda campaigns integrate digital militia into cohesive networks and direct them to harass and threaten specific targets, and the massive levels of coordination involved alters the scale and nature of the attacks.

Existing studies of online hate speech using machine learning approaches have produced powerful sociolinguistic insights into the content of online hate speech. However, they have struggled to keep pace with the fast-changing nature of online speech and hate speech that is implicit or coded.[62] They usually rely on standardized lists of hate speech (e.g., Hatebase.org), although a few recent studies have developed a more sophisticated categorization of different linguistic variants of hate speech.[63] The methodological requirements of quantitative studies mitigate against a fine-grained and culturally-informed analysis of the various types of online hate speech, including anti-human rights speech.

Quantitative studies of the causal effects of online hate speech necessarily rely on visible outcomes such as hate crimes and lethal acts of violence. Although international organizations such as Frontline Defenders track data on non-lethal harms against defenders such as threats, criminal arrests, and detentions, these have not been analyzed and substantiated in quantitative social science studies. As noted earlier, the criminalization of human rights activism is a widespread strategy of illiberal governments around the world, and UN Special Rapporteur Mary Lawlor observes that many killings of defenders are preceded by threats and criminalization.[64] The

---

[61] The exception being the 2017 genocide in Myanmar that was incited and coordinated on Facebook by the Burmese military (Tatmadaw); *see* Paul Mozur, *A Genocide Incited on Facebook, with Posts from Myanmar's Military*, N.Y. TIMES (Oct. 15, 2018), https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html; Alexandra Stevenson, *Facebook Admits It Was Used to Incite Violence in Myanmar*, N.Y. TIMES (Nov. 6, 2018) https://www.nytimes.com/2018/11/06/technology/myanmar-facebook.html.

[62] Mai ElSherief et al., *Hate Lingo: A Target-Based Linguistic Analysis of Hate Speech in Social Media,* ICWSM (2018a); Mai ElSherief et al., *Peer to Peer Hate: Hate Speech Instigators and Their Targets*, ICWSM (2018b).

[63] ElSherief et al., *Hate Lingo, supra* note 62.

[64] Final Warning: Death Threats and Killings of Human Rights Defenders: *Report of the*

use of social media to promote "lawfare"[65] compels us to extend the insights of social science inquiries into the causal effects of hate speech to also examine digital authoritarianism.

This entails an inquiry into the broader consequences of online hate speech and its effects on social norms and political institutions. Campaigns undertaken by powerful actors may contribute to a climate of intolerance, impunity, and corruption by eroding social norms against threatening speech, and fraying bonds of trust and cooperation within and between societal groups.[66] They may undermine a population's commitment to fundamental democratic norms such as human rights and the rule of law, and exert a chilling effect on inquiries into corruption or violations of human rights.[67] The suppression of human rights discourse may enable further harms such as criminalization or physical violence. There is still much work to be done in documenting the conditioning effects of speech that set the stage for physical attacks by preparing a population to accept violence.[68] Here, speech is less a proximate cause than a preparatory act that contributes to the early phases of a causal sequence that may culminate in violence.

Finally, there has been little systematic analysis of the hidden harms experienced by human rights defenders that are not tracked in government or civil society statistics. To this end, I conducted in-depth interviews with human rights defenders in Colombia and Guatemala, about the effects of online hate speech on them personally and the context-specific ways that online harassment affects human rights work in their country. Employing a qualitative study of human agency enmeshed in hermeneutic "webs of meaning,"[69] facilitates the study of consequences that are seldom visible in the statistics: psychological harms, self-censorship, burnout, and broader political outcomes such as the undermining of trust in the knowledge claims

---

Special Rapporteur on the Situation of Human Rights Defenders, Mary Lawlor, Human Rights Council, 46 U.N. GOAR, U.N. Doc. A/HRC/46/35 (Dec. 24, 2020), p. 5.

[65] Understood as the misuse of legal means for political or military ends. *See* Brooke Goldstein, *Lawfare: Real Threat or Illusion, Address Before The Princeton Club* (Nov. 5, 2010) https://web.archive.org/web/20160112224349/http:/www.thelawfareproject.org/Articles-by-LP-Staff/lawfare-real-threat-or-illusion.html#_ftn2.

[66] Michael Bang Petersen, *The Evolutionary Psychology of Mass Mobilization: How Disinformation and Demagogues Coordinate Rather than Manipulate*, 35 CURR. OPINION IN PSYCH. 71 (2020).

[67] *See* DANIELLE K. CITRON, HATE CRIMES IN CYBERSPACE 196-98 (2014) (discussing the "devastating impact" and chilling effects of online harassment).

[68] Molly K. Land & Rebecca Hamilton, *Chapter 6: Beyond Takedown: Expanding the Tool Kit for Responding to Online Hate*, *in* PROPAGANDA AND INTERNATIONAL CRIMINAL LAW: FROM COGNITION TO CRIMINALITY 143 (Predrag Dojčinović ed., 2020).

[69] MARK BEVIR & JASON BLAKELY, INTERPRETATIVE SOCIAL SCIENCE: AN ANTI-NATURALIST APPROACH 1 (2018).

of human rights activists and organizations. More broadly, the harmful effects of hate speech are extensively documented and include emotional duress, poor health, and diminished educational outcomes for individuals who are targeted.[70]

First, I conducted eighty-one semi-structured interviews;[71] fifty-six with human rights defenders (thirty-nine from Colombia and Guatemala, and seventeen from Ireland, Nigeria, the Philippines, Serbia, the United Kingdom, and the United States); twelve with journalists from Colombia, Guatemala, the Philippines, and the United States; ten with representatives of national governments of Colombia or Guatemala, or international agencies such the United Nations or Inter-American Commission working in these two countries; and three with academic experts on social media. Additionally, there were regular informal conversations with eleven staff from Facebook/Meta and Twitter about their hate speech and content moderation policies.[72]

Second, due to the fragmentation and incompleteness of existing data on attacks on defenders, I created a database of killings of hundreds of defenders in Colombia and Guatemala in 2020 that included their names, the date and place of their killing, the identity of the perpetrator (if known), and the corroborating source of information.[73] This helped to determine where the killings of defenders were happening in the country and identify any geographic patterns and relation to social media coverage.

Finally, I created a database of anti-human rights speech online by collecting and hand coding 400 Twitter posts targeting defenders in Colombia and Guatemala.[74] The categories of anti-human rights speech used in the coding draw upon forms of speech and types of posts identified by human

---

[70] In Harper v. Poway, the Ninth Circuit Court of Appeals cited seven academic studies on emotional harms of derogatory speech to justify upholding the right of a school district to prevent a student from wearing a homophobic t-shirt in the classroom. Harper v. Poway Unified School Dist., 445 F.3d 1166, 1179 (9th Cir. 2006). *See* Koustuv Saha et al., *Prevalence and Psychological Effects of Hateful Speech in Online College Communities*, PROC. ACM WEB SCI. CONF. 255 (2019) (describing the harmful psychological effects of online hate speech); Martin H. Teicher, *Hurtful Words: Association of Exposure to Peer Verbal Abuse with Elevated Psychiatric Symptom Scores and Corpus Callosum Abnormalities*, 168 AM. J. PSYCHIATRY 213 (2010) (outlining long-term psychological harms associated with verbal abuse).

[71] Using a snowball sampling method to identify potential interviewees, the majority of whom were based in urban areas (86 percent) and female (55 percent).

[72] In 2020, Facebook requested my input on their developing policies on implicit hate speech and COVID-19-related hate speech. These consultations were unremunerated.

[73] The list of sources included United Nations agencies, official government statistics, reports from human rights organizations, and local press outlets in each country. This database is available upon request.

[74] Database and coding scheme are available from author upon request.

rights defenders. The interactions between the different categories of anti-human rights speech are also of interest since they help us to understand what forms of speech frequently appear together and may have interactive effects.

IV. DIGITAL AUTHORITARIANISM IN COLOMBIA AND GUATEMALA

In order to grasp the effects of digital authoritarianism, we have to understand the historical and political contexts in which human rights defenders operate. Digital authoritarianism does not operate in a vacuum; online propaganda accompanies real-world anti-democratic practices such as the killing of defenders and the criminalization of human rights activism.

Colombia and Guatemala are appropriate contexts for developing a granular understanding of state information operations and their consequences for human rights work. Both countries had decades-long armed conflicts driven by land inequality that resulted in the deaths of over 200,000 citizens and high levels of political violence continue.[75] The United States played a significant role in providing military support and training for successive Colombian and Guatemalan governments, including military dictatorships.[76] In both countries, well-organized human rights movements emerged during the armed conflicts and experienced violent repression by state security services and paramilitaries.[77]

A hallmark of authoritarianism is the dismantling of the administrative state,[78] and, the accountability mechanisms that investigate and prosecute corruption and politically motivated crimes by state officials.[79]

---

[75] Alejandro Castillejo, *La Escala Humana de la Herida: Apropiaciones y Traducciones del Daño en Colombia, in* COLOMBIA CONTEMPORÁNEA: MIRADAS DISCIPLINARES DIVERSAS (Mauricio Nieto Olarte ed., 2017); GREG GRANDIN, THE LAST COLONIAL MASSACRE: LATIN AMERICA IN THE COLD WAR (2011); GONZALO SÁNCHEZ & RECARDO PEÑARANDA PASADO Y PRESENTE DE LA VIOLENCIA EN COLOMBIA (2007); RICHARD ASHBY WILSON, MAYA RESURGENCE IN GUATEMALA: Q'EQCHI' EXPERIENCES (1999).

[76] *See* GUATEMALA MEMORY OF SILENCE: REPORT OF THE COMMISSION FOR HISTORICAL CLARIFICATION CONCLUSIONS AND RECOMMENDATIONS 19-20 (1999) (describing the role of the United States in Guatemala); Julio Ramirez Montañez, *Fifteen Years of Plan Colombia The Recovery of a Weak State and the Submission of Narco-Terrorist Groups?* 7 ANALECTA POLIT. 315 (2017) (describing the United States in the Colombian conflict).

[77] Cora Currier & Danielle Mackey, *The Rise of the Net Center: How an Army of Trolls Protects Guatemala's Corrupt Elite*, THE INTERCEPT (Aug. 27, 2017), https://perma.cc/VFU5-4YSK.

[78] Gillian E. Metzger, *1930s Redux: The Administrative State Under Siege*, 131 HARV. L. REV. 1, 3 (2017) (noting the attack on administrative governance).

[79] Samuel Issacharoff, *The Corruption of Popular Sovereignty* (2020) (NYU School of Law, Public Law Research, Working Paper No. 20-02) (available on SSRN); Nadia Urbinati, *Political Theory of Populism*, 22 ANNUAL REV. POL. SCI. 111 (2019).

In Colombia and Guatemala, human rights activists are at the forefront of anti-corruption efforts, for instance, by supporting the UN's International Commission Against Impunity in Guatemala (CICIG). They have also pushed for criminal accountability for conflict-era mass atrocities committed by right-wing paramilitaries and the army.[80] These human rights campaigns have exposed defenders to violent repression at the hands of state and private actors, and after the 2016 Peace Accords in Colombia, the number of killings of human rights defenders rose sharply from fifty-six[81] to ninety-two;[82] an increase of 64 percent. Defenders working in rural areas are regularly at the highest risk of violence. According to the database created for this study, in 2020 in Colombia, 81 percent of defenders killed were in rural areas, and in Guatemala in the same year, 100 percent of defenders killed were in rural areas.[83] Neither country has an effective criminal justice system for investigating homicides. The 2018 impunity rates are the same in Colombia[84] and Guatemala[85]; 98 percent. In Colombia, the impunity rate for murders of defenders has historically stood at 95 percent.[86]

Anti-human rights speech online is remarkably similar in Colombia and Guatemala and invokes a Cold War narrative of the patriotic defense of the nation from foreign, Marxist, and terrorist destabilizers who are common criminals, not legitimate political representatives.[87] In both countries, governments have sought to mobilize their base online against human rights activists and opposition figures.[88] Revelations have surfaced repeatedly that Colombian military intelligence surveils and hacks the devices of opposition political and civil society figures,[89] and the president's office orchestrates a

---

[80] Such as the massacres at Mapiripán in Colombia, and Dos Erres and Rio Negro in Guatemala.

[81] *See* FRONT LINE DEFENDERS, GLOBAL ANALYSIS 2016 (2017), https://www.frontlinedefenders.org/en/resource-publication/2016-annual-report.

[82] *See* FRONT LINE DEFENDERS, GLOBAL ANALYSIS 2017 (2018), https://www.frontlinedefenders.org/en/resource-publication/annual-report-human-rights-defenders-risk-2017.

[83] In Colombia in 2020, 247 out of 304 defenders killed were in rural areas (81 percent). There were twenty-seven defenders killed in urban areas (9 percent). We were unable to find reliable sources on the location of the remaining 10 percent. In Guatemala, 100 percent of the fifteen defenders killed in 2020 were based in rural areas.

[84] *See* Parker Asman, *Guatemala Impunity Report Shows Limits of Anti-Graft Body*, INSIGHT CRIME (June 18, 2019), https://insightcrime.org/news/brief/guatemala-impunity-report-limits-anti-graft-body/; Michel Forst, *Report of the Special Rapporteur on the Situation of Human Rights Defenders*, U.N. Doc. A/HRC/43/51/Add.1 (Dec. 26, 2019).

[85] Asman, *supra* note 84.

[86] *See* Forst, *supra* note 84 (figures are not available for Guatemala).

[87] ABBAS ET AL., *supra* note 6; Gregory, *supra* note 23.

[88] Currier & Mackey, *supra* note 77.

[89] Redacción Judicial, *Las "Carpetas Secretas" de Inteligencia Militar: ¿A Quiénes Iban Dirigidas y Para qué?*, EL ESPECTADOR (May 3, 2020 9:56 PM),

social media campaign to slander opponents.[90] In Guatemala, military intelligence deploys sophisticated cellphone-hacking and surveillance software such as Pegasus against human rights activists.[91] The leading anti-human rights account in Guatemala from 2016 to 2020 (@LordVaderGT) was reportedly operated by an associate of former Vice-President Felipe Alejos Lorenzana, who was sanctioned in 2020 by the Department of State for corruption under Section 7031(c).[92] Furthermore, there is ample evidence that leading Colombian and Guatemalan anti-human rights figures coordinate with one another.[93]

Human rights work is frequently criminalized in both countries. For instance, eight human rights defenders were arrested in November 2018 in San Luis de Palenque, Colombia after protesting the presence of contractors from the Canadian energy company Frontera Energy on their private lands.[94] Frontera Energy privately contracted the Colombian army to protect its activities, lodged a criminal complaint, and the defenders were arrested for criminal conspiracy, violence against a public servant, and obstruction of public roads.[95] Two were charged with attempted homicide in connection with their leadership of protests during 2016 and 2018 in response to the failure of Frontera Energy to compensate communities affected by environmental damage.[96] The defenders were beaten, arrested, and detained

---

https://www.elespectador.com/judicial/las-carpetas-secretas-de-inteligencia-militar-a-quienes-iban-dirigidas-y-para-que-article-917751/.

[90] La Liga Contra el Silencio, *En las Entrañas de una "Bodeguita" Uribista*, EL ESPECTADOR, (Feb. 6 2020), https://www.elespectador.com/politica/en-las-entranas-de-una-bodeguita-uribista-article-903239/.

[91] *See* Ángel Sas & Coralia Orantes, *Espionaje Ilegal del Gobierno: Aquí Está la Investigación de Nuestro Diario (Parte I)*, NÓMADA (Aug. 6, 2018), https://nomada.gt/pais/la-corrupcion-no-es-normal/espionaje-ilegal-del-gobierno-aqui-esta-la-investigacion-de-nuestro-diario-parte-i/?utm_source=nomada_ux&utm_medium=hay_mas_autor (describing the use of Pegasus by Guatemalan security services).

[92] Press Release, US Dept. of State, Public Designation of Current and Former Members of the Guatemalan Congress Due to Involvement in Significant Corruption (Oct. 28, 2020), https://gt.usembassy.gov/public-designation-of-current-and-former-members-of-the-guatemalan-congress-due-to-involvement-in-significant-corruption/.

[93] *See e.g.*, Guatemalan Ricardo Mendez Ruiz's tweet of his meeting with former Colombian President Álvaro Uribe; Ricardo Mendez Ruiz (@RMendezRuiz), TWITTER (Oct. 17, 2018 6:35 PM), https://twitter.com/RMendezRuiz/status/1052689655353626624?s=09.

[94] *See* Forst, *supra* note 84; Edinson Arley Bolaños, *La Detención de Líderes Sociales que llega a las Naciones Unidas*, EL ESPECTADOR (Mar. 4, 2020 6:00 AM), https://www.elespectador.com/colombia-20/conflicto/la-detencion-de-lideres-sociales-que-llega-a-las-naciones-unidas-article/.

[95] Forst, *supra* note 84, ¶ 29.

[96] *Id.*

by the police, and the charges have yet to be proven in a court of law.[97] This is part of a wider pattern and 202 Colombian defenders protecting environmental rights have been prosecuted since 2012.[98] The criminalization of human rights activists is a practice that is reinforced daily by social media messaging, as we will see in detail in Part V.

In Guatemala between 2016 and 2020 the Myrna Mack Foundation documented 323 criminal complaints against fifty-nine defenders.[99] In September 2019, the Guatemalan human rights organization, La Unidad de Protección de Defensoras y Defensores de Derechos Humanos-Guatemala (UDEFEGUA), reported ninety-one outstanding arrest warrants for defenders in the department of Huehuetenango and fifty-two in Alta Verapaz.[100] Frivolous litigation is facilitated by the fact that private citizens can lodge a criminal complaint (*denuncia*) against any person for vaguely-formulated crimes such as conspiracy, abuse of authority, violation of the Constitution, revealing confidential information, sedition, trespass, defamation, and "illicit association."[101] According to a UN official in Guatemala, 30 percent of arrest warrants issued against defenders are for trespass.[102] Of the 323 *denuncias* cited above, fifty-five were initiated by the pro-military "Foundation Against Terrorism" and the majority of these were simultaneously released on Facebook and Twitter.[103]

When powerful political or economic actors bring *denuncias*, they are more likely to lead to prosecutorial investigations, indictments, and warrants of arrest. Private complaints have resulted in defenders being arrested and detained without trial for years, as in the cases of Guatemalan indigenous rights activists Daniel Pascual and Abelino Chub Caal.[104] Even if a criminal complaint against a defender is dismissed as frivolous, the defender may be detained, jailed, face high legal defense costs, and significant disruption to their work. In some instances, lawfare tactics force defenders to leave the country. Former Guatemalan attorneys-general Thelma Aldana and Claudia Paz y Paz fled Guatemala after a judge issued warrants for their arrest on spurious charges and Aldana later received asylum in the United States.[105]

---

[97] *Id.*

[98] *Id.*

[99] Fundación Myrna Mack, *Red De Impunidad: Persecución Mediática y Jurídica*, 66 (2020).

[100] Interview with Guatemalan human rights defender (2019).

[101] Fundación Myrna Mack, *supra* note 99, at 70.

[102] Interview with UN official (2019).

[103] Fundación Myrna Mack, *supra* note 99, at 67-68.

[104] *Two Years in Jail for Protecting his Community's Land in Guatemala,* INDEPENDENT CATHOLIC NEWS (Jul. 17, 2021), https://www.indcatholicnews.com/news/42649.

[105] *Ex-Guatemala Prosecutor Granted Asylum in U.S.*, AP NEWS (Feb 24, 2020), https://apnews.com/article/ce4c035ff39ba0f0b2362adaa529194e.

The differences between the countries are also generative of comparison. Colombia has nearly three times the population of Guatemala, and its government possesses greater institutional capacity, both militarily and otherwise.[106] The Colombian state's coercive power is limited, however, to certain areas of the country, and observers have referred to it as a "fragmented state" in which different armed actors exert "oligopolies of coercion" over the territories they control.[107] According to CICIG, the Guatemalan state has been "captured" by organized crime and elements of the security apparatus.[108] In both countries, the state's involvement in organized crime is centered on illegal drugs.[109] However, Colombia is a net narcotics producer and exporter whereas Guatemala is positioned as a major transit country for illegal drugs destined for the United States.[110]

The two countries also have markedly different histories and cultures. Due to its proximity to the United States and smaller size, Guatemala has experienced greater US influence in its politics, including a CIA-sponsored coup in 1954 that replaced reformist president Jacobo Árbenz with an anti-communist military dictatorship.[111] Colombia's indigenous population is less than five percent of the total population, whereas the 2018 Guatemalan census found that 43 percent of the population identify as indigenous, one of the highest percentages in Latin America.[112] A United Nations-sponsored truth commission found that in the 1980s, the military regime of Rios Montt

---

[106] *Colombia Population*, WORLDOMETER (2021), https://www.worldometers.info/world-population/colombia-population/.

[107] Gustavo Duncan, *Drug Trafficking and Political Power: Oligopolies of Coercion in Colombia and Mexico*, 41 LATIN AM. PERSPECTIVES 18 (2013).

[108] UNITED NATIONS COMMISSION AGAINST IMPUNITY IN GUATEMALA (CICG), GUATEMALA: UN ESTADO CAPTURADO (2019).

[109] A Colombian government internal investigation "Operation Baton" found that army generals had business links with Mexican drug traffickers and sold information to the FARC rebels. Unidad Investigativa, *Operación Bastón: el Destape de la Corrupción en el Ejército*, EL TIEMPO (May 17, 2020), https://www.eltiempo.com/unidad-investigativa/operacion-baston-que-revela-corrupcion-dentro-del-ejercito-496292. A 2019 Amnesty International Report identified sixty criminal structures in the highest levels of the Guatemalan state between 2007-2018: AMNISTÍA INTERNACIONAL, ÚLTIMA OPORTUNIDAD DE JUSTICIA. PELIGROSOS RETROCESOS PARA LOS DERECHOS HUMANOS Y LA LUCHA CONTRA LA IMPUNIDAD EN GUATEMALA 4 (2019).

[110] UNITED STATES DEPARTMENT OF STATE BUREAU OF INTERNATIONAL NARCOTICS AND LAW ENFORCEMENT AFFAIRS, INTERNATIONAL NARCOTICS CONTROL STRATEGY REPORT VOL. 1, at 113, 141 (2021).

[111] UN COMMISSION FOR HISTORICAL CLARIFICATION, *supra* note 76, at 19.

[112] DANE: POBLACIÓN INDÍGENA DE COLOMBIA RESULTADOS DEL CENSO NACIONAL DE POBLACIÓN Y VIVIENDA 2018 (2019), https://www.dane.gov.co/files/investigaciones/boletines/grupos-etnicos/presentacion-grupos-etnicos-2019.pdf; Silvel Elías, *Indigenous World 2020: Guatemala*, IWGIA (2021), https://www.iwgia.org/en/guatemala/3622-iw-2020-guatemala.html.

committed genocide against the Maya-Ixíl group.[113] Whereas the Colombian armed conflict ended in 2016, the Guatemalan peace accords were signed twenty years earlier in 1996, permitting a comparison of the effects on political discourse of temporal distance from an armed conflict.

V. THE CONTENT OF ANTI-HUMAN RIGHTS SPEECH

The empirical part of this study began by interviewing Colombian and Guatemalan defenders about the most frequent tropes of anti-human rights speech. In their interviews, human rights defenders reported being subjected to smears online that refer to them as "communists," "guerrillas," "criminals," and "terrorists" bent on destroying the state.[114] They are labelled narcotics traffickers, and as having undesirable personal characteristics such as being "disgusting," "corrupt," and "violent." They are accused of treason and being in the pay of foreign actors such as George Soros. Racism, misogyny, and anti-LGBTQ+ slurs are common. They are regularly called, in typically authoritarian language, "enemies of the state" or "the internal enemy."[115] Death threats that would normally be removed by social media platforms in the United States and Western Europe often circulate unchecked in Latin America; for instance, the "Black Eagles," a group representing paramilitaries involved in the narcotics trade in Colombia, posts death lists online that include prominent human rights defenders who they call "guerrillas in disguise" (*guerrilleros camuflados*).[116]

In order to better understand the content of anti-human rights speech online, this study collected and hand coded 400 Twitter posts (200 per country) between December 2018 and December 2020 in a convenience sample that included relevant keywords and hashtags and the posts of prominent defenders and anti-human rights accounts in each country.[117] Consistent with grounded theory in social science, and drawing on the

---

[113] UN COMMISSION FOR HISTORICAL CLARIFICATION, *supra* note 76, at 38-39.

[114] *See* Forst, *supra* note 84, ¶ 27.

[115] *See* CRIMINALIZACIÓN, ATAQUES MEDIÁTICOS Y DISCURSO DE ODIO, FUNDACIÓN MYRNA MACK 4 (2020) (describing how defenders are termed the "internal enemy" in Guatemala).

[116] Colombia: Águilas Negras, INSIGHT CRIME (Mar. 9, 2017), https://insightcrime.org/colombia-organized-crime-news/aguilas-negras/ (analyzing the Black Eagles); *see* @ONIC_Columbia, TWITTER (Mar. 11, 2020 3:24 PM), https://twitter.com/ONIC_Colombia/status/1237821770083840001.

[117] All posts were double-blind coded in the twelve categories by two trained coders, and then reconciled by the author with both coders.

existing social science literature on dehumanization,[118] stigmatization,[119] revenge,[120] and threatening speech,[121] I identified twelve distinct categories of anti-human rights posts. Together, they comprise the overarching category of "anti-human rights speech."

1. *Direct threats of harm:* direct calls to kill or injure an individual or their family, posting a home address, referring to a death squad, or images of harm, violence, or death.

2. *Implied threats of harm:* non-specific calls for something to be done, wishing harm would befall them, negative statements about life expectancy, or images indicating the above.

3. *Accusations of corruption:* direct or implied accusations that a person or an organization is corrupt or is engaged in fraudulent activities.

4. *Accusations of subversion and terrorism:* claims that the person is a communist, Marxist, terrorist, guerrilla, assassin, or images indicating the above.

5. *Assertions of anti-patriotic behavior:* statements that the target is a traitor, betraying the country or way of life, an enemy of the people, or is serving foreign interests.

6. *Accusations of criminality:* statements that the person is a criminal, delinquent, bandit, fugitive of justice, part of conspiracy or criminal network, structure or organization, or calls to charge or jail them.

7. *Surveillance:* photos or videos taken in public place without the knowledge or permission of the target.

8. *Doxing:* non-consensual release of private or identifying information, including documents, private images, or other private materials, with the intention of harassing, shaming, or inflicting harm.

9. *Dehumanization:* statements or images that the target is non-human, including an animal, a virus, or a non-human object.

---

[118] Emile Bruneau, et al., *Denying Humanity: The Distinct Neural Correlates of Blatant Dehumanization.* 147 J. EXPERIMENTAL PSYCH.: GENERAL 1078 (2018); Bernhard Leidner et al., *Ingroup Glorification, Moral Disengagement, and Justice in the Context of Collective Violence*, 36 PERSONALITY & SOC. PSYCH. BULL. 1115 (2010).

[119] ERVING GOFFMAN, STIGMA: NOTES ON THE MANAGEMENT OF SPOILED IDENTITY (1963); Mark L. Hatzenbuehler, Susan Nolen-Hoeksema & John Dovidio, *How Does Stigma "Get Under the Skin"?: The Mediating Role of Emotion Regulation*, 20 PSYCH. SCI. 1282 (2009).

[120] Joshua Conrad Jackson, Virginia K. Choi & Michele J. Gelfand, *Revenge: A Multilevel Review and Synthesis*, 70 ANN. REV. PSYCH. 319 (2019).

[121] Ángel Gómez et al., *Responses to Endorsement of Commonality by Ingroup and Outgroup Members: The Roles of Group Representation and threat*, 39 PERSONALITY & SOC. PSYCH. BULL. 419 (2013).

10. *Gender or sexuality-based disparagement:* statements that the person is LGBT+, or questioning a person's gender or sexuality, or accusations of sexual perversion.

11. *Narratives from the armed conflict:* conflict-era slur or denial of documented massacre, mass atrocity, or other crime during armed conflict.

12. *Stigmatization:* insults based on race, ethnicity, national origin, or statements that the target is disgusting or offensive, or accusations of mental illness or substance abuse, or images or emojis conveying the above.

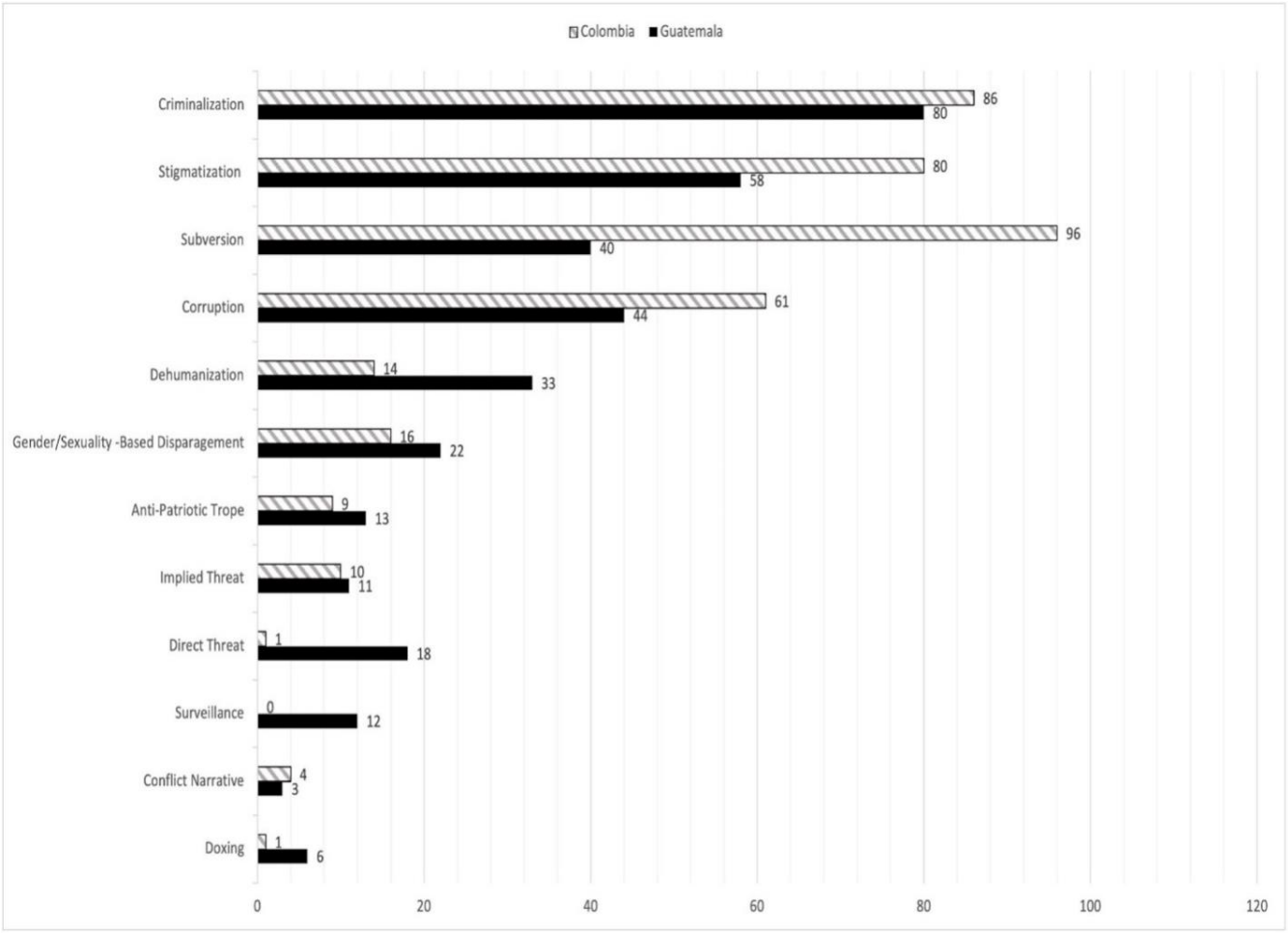

Figure 1: Categorization of 400 Twitter Posts in Colombia and Guatemala in 2020

The most frequent categories of anti-human rights speech in Colombia are 1. Subversion; 2. Criminalization; 3. Stigmatization; and 4. Corruption. In Guatemala, the main categories of anti-human rights speech are; 1. Criminalization; 2. Stigmatization; 3. Corruption; and 4. Subversion. During coding, it became apparent that a majority of posts contained more than one category of anti-human rights speech. Therefore, I coded all the types present in a single post and analyzed the interactions between categories of speech occurring together.[122] The visualizations in Figures 3 and 4 illustrate the number of times that categories occur together in a single post. They represent a preliminary Principal Components Analysis that highlights terms that cluster together and reduces the complexity of a data set consisting of many interrelated variables to their core attributes.[123]

[122] In R, we processed the column with the labels for each post by separating out each label into its own column. We then checked whether there were two or more labels assigned to a given post. If they were, we took the "n choose 2" combinations of these labels, which were not directional. After iterating over every post for both Guatemala (199 posts) and Colombia (157 posts) separately, we counted the number of times a given combination had occurred in among the posts and used the "graph" package in R to visualize the 2-way connections among labels. The width of the bar connecting the labels indicates the strength of the connection.

[123] IAN. T. JOLLIFFE, PRINCIPAL COMPONENTS ANALYSIS (2d ed., 2002).

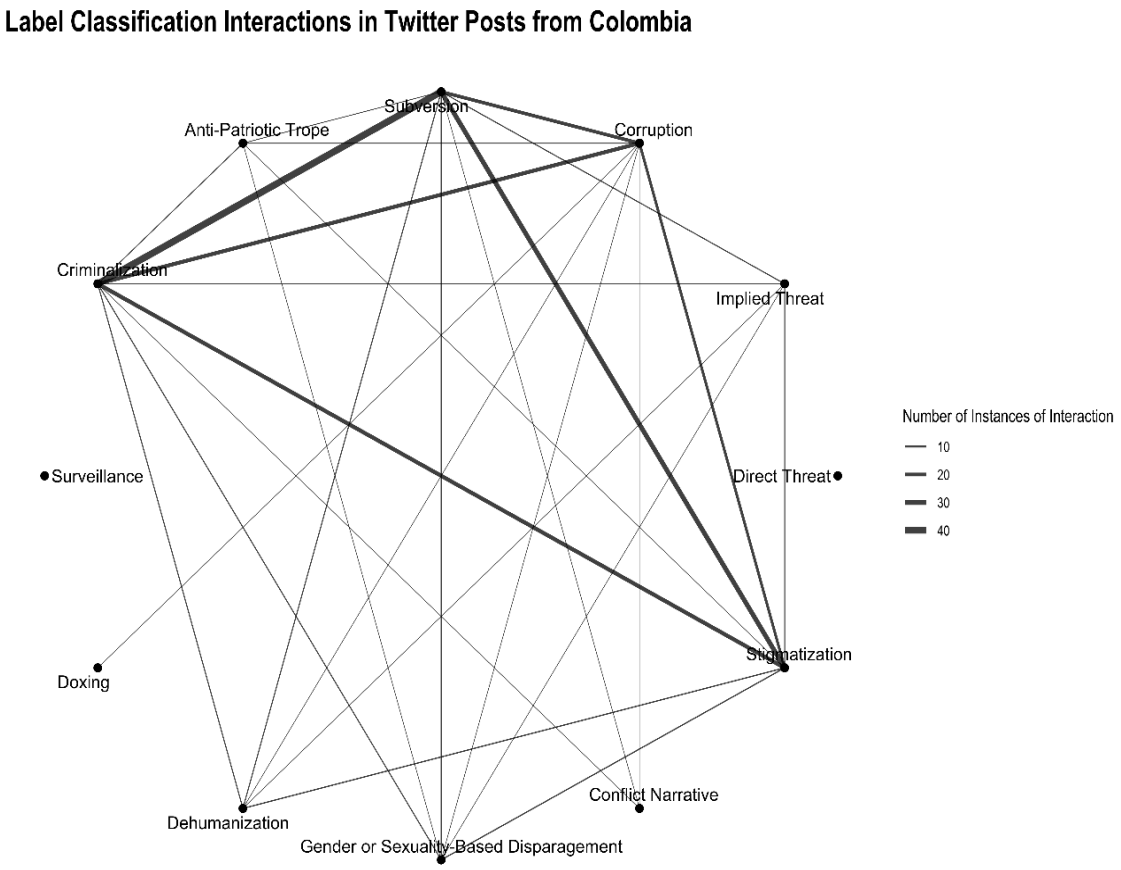

Figure 2: Colombia: interaction of categories of anti-human rights speech

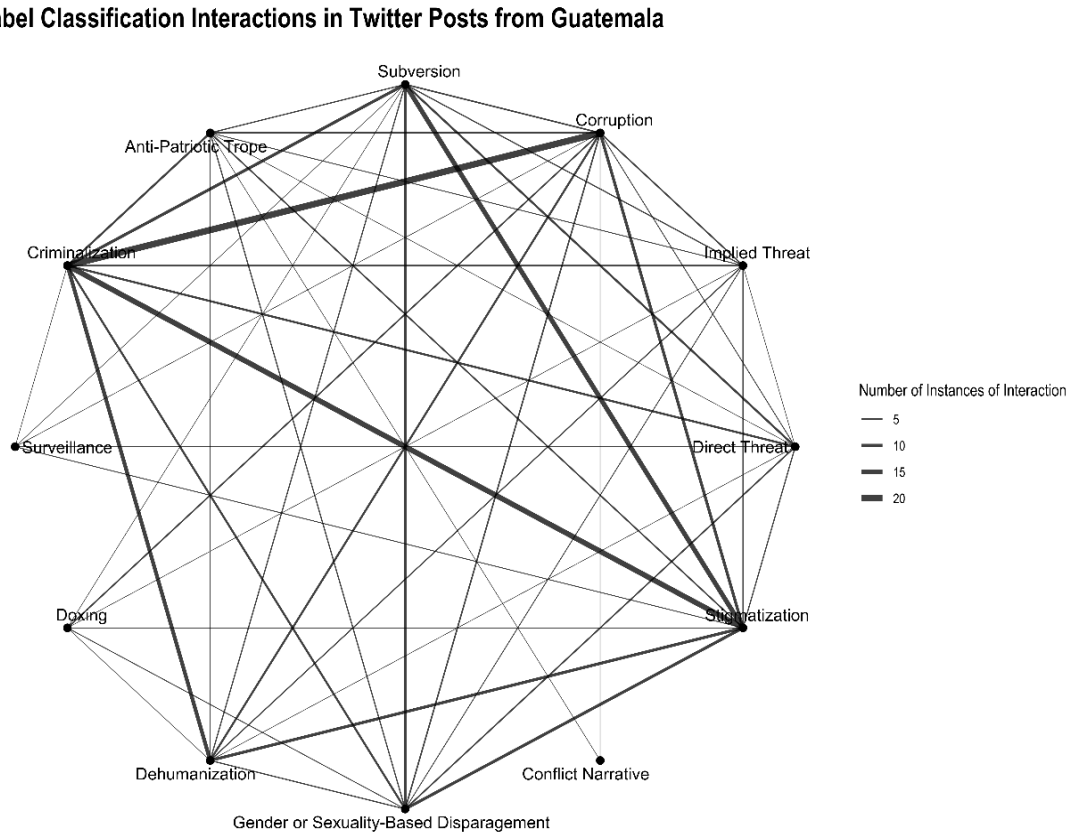

Figure 3: Guatemala: interaction of categories of anti-human rights speech

These empirical results support several findings: First, the content of anti-human rights speech in Colombia and Guatemala is remarkably similar and stable and accuses human rights defenders of being corrupt, criminals, and subversives and stigmatizes them. Methodologically, the results also demonstrate the value of a fine-grained analysis of the content of online hate speech. Social science studies in the Minority Model have generally used binary categories, such as Relia et al.'s hate speech/not-hate speech,[124] and Müller and Schwarz's anti-immigrant speech/not anti-immigrant speech.[125] If we go beyond binary approaches to online posts, then we stand a better chance of grasping the discursive nature of anti-human rights speech.

Second, anti-human rights speech is also highly contextual and

[124] Relia, *supra* note 56.
[125] Müller & Schwarz, *supra* note 19.

culturally specific. Digital authoritarianism commonly uses coded or contextually specific speech to evade the content moderation processes of the platforms. For instance, pro-military social media accounts in Guatemala refer to the van used for surveillance operations as the "ice cream truck" (*carrito de helados*).[126] This expression resonates with Guatemalan activists because of the ice cream truck's association with military intelligence and death squads in the 1980s. Yet this term is benign in the Colombian context; it means literally, "ice cream truck." Surveillance, including posting photos of targets in public settings such as restaurants in real time, is prevalent in Guatemala but absent in Colombia. Likewise, in Colombia, the statement that someone will "send the motorcycle" (*manda la moto*) is commonly understood as a death threat since armed assassins in Colombia often arrive on a motorcycle. The statement is not considered threatening in Guatemala where the practice and expression are not prevalent.

Third, the interactions between categories of anti-human rights speech are pronounced and may be meaningful, although we do not yet fully understand their significance. In Guatemala, the most frequent combinations of anti-human rights speech are Corruption-Criminalization, and Stigmatization-Criminalization-Corruption. The connections between the Criminalization-Subversion-Stigmatization in Colombia are more concentrated than in Guatemala. This suggests that in future studies, rather than examining one category of speech (such as incitement, threat, or dehumanization), we might instead investigate whether certain combinations of online speech are jointly sufficient to cause offline harms.

Finally, and perhaps most importantly for this article, the number of direct and implicit threats including incitement to violence are conspicuously few. In Colombia, there are fewer threatening posts than in Guatemala, even though the number of defenders killed in Colombia in 2020 (304) was twenty times higher than Guatemala (15). This observation is supported by further analysis of the offline effects in Guatemala of online speech. I conducted a time-series regression using the database of killings of defenders in 2020 and data of threatening posts collected on social media and found no statistically significant correlation between the two.

Therefore, an initial analysis suggests that the frequency of online threats of violence does not correlate with actual levels of violence. Of course, the research design could have been flawed. Methodologically, the relatively low number of threatening posts could be related to the

---

[126] *See* Ángel Sas & Coralia Orantes, *A ellos Espiaba el Gobierno con un Carrito de Helados*, NÓMADA (Aug. 7, 2018), https://nomada.gt/pais/actualidad/a-ellos-espiaba-el-gobierno-con-un-carrito-de-helados-la-investigacion-de-nuestro-diario-parte-i/ (discussing the 'ice cream truck' in government surveillance); *see e.g.* @LordVaderGT TWITTER (Dec. 24, 2020), https://twitter.com/LordVaderGT/status/1342176983569068032?s=19.

convenience sample, and here, it is worth noting that a quantitative analysis of attacks on specific occupational groups such as human rights defenders or journalists is impeded by the relatively low number of posts overall compared with, say, the 500,000 posts and comments on the anti-immigrant Alternative for Germany Facebook page in 2016. Additionally, social media speech is a rapidly moving target as platforms have adopted increasingly aggressive content moderation measures in recent years and by one estimation now remove approximately 62 percent of hate speech flagged by users.[127] During the data collection phase of this study, it was apparent that as Twitter's content moderation removed more explicit threats, the discourse of state-aligned accounts shifted to harassment and denigration.

These findings compel us to question whether the Minority Model's emphasis on identifying a causal relationship between online speech and offline harms is the most appropriate paradigm for comprehending the range of outcomes of digital authoritarianism. If it is not, then we may have to employ different methodological techniques to access a broader range of negative outcomes of online speech.

VI. THE HIDDEN HARMS OF ANTI-HUMAN RIGHTS SPEECH

The findings of the last section with respect to direct online threats and their lethal consequences suggest that we might consider in more detail the non-lethal effects of online speech (such as intimidation and silencing) on defenders that have been emphasized in the Political Model. This section summarizes the evidence from qualitative interviews with human rights defenders, journalists, and UN and government officials in Colombia and Guatemala. Human rights defenders reported a range of negative effects of online anti-human rights speech on their professional and personal lives, including:

A. Fear and Intimidation

An overwhelming majority (92 percent) of Colombian and Guatemalan defenders interviewed in this study reported experiencing fear and being intimidated when targeted online. One defender explained, "They use social media to gain influence and to manipulate your feelings and your psyche. They make threats to sew chaos . . . because that's how they control you. . . . These are psychological operations to make you paranoid and they use fear to control you. It's Propaganda 2.0."[128] A journalist reflected that "it causes

[127] DIDIER REYNDERS (COMMISSIONER FOR JUSTICE), COUNTERING ILLEGAL HATE SPEECH ONLINE: 6TH EVALUATION OF THE CODE OF CONDUCT, EUROPEAN COMMISSION (Oct. 7, 2021).

[128] Interview with Guatemalan human rights defender (2018).

personal instability to live under such pressure," which he saw as a conscious governmental strategy of "psychological counterinsurgency."[129] Defenders underscored how the extraordinary reach and immediacy (or "virality") of social media is qualitatively different from traditional media. As a result, threats feel more personal and invasive of their privacy for some defenders, as a female indigenous rights activist from Guatemala explained, "The harassment and death threats on social media began when I denounced a massacre by the army in my hometown. . . . I got scared about my security and my family's security…they profiled my entire family on social media. . . . Hate speech online is damaging in a way that is very personal and intimate. The threat is on your phone that is right there next to you."[130] When posts reveal their home address, identify family members, or contain surveillance photos of defenders at restaurants, protests, or public meetings, defenders describe heightened levels of fear. The scope and capacity of surveillance operations is most apparent when intimate pictures of defenders in foreign countries appear online. For example, photos of former Guatemalan attorney-general Thelma Aldana using a public bathroom in Orlando circulated on Twitter in December 2019, and the ultra-rightwing Colombian vigilante Andrés Escobar posted a photo on Twitter of exiled human rights activist Beto Coral walking with his family in a New York City park in July 2021.[131]

B. Reputational Damage

Of the defenders interviewed, 90 percent reported that online attacks had damaged their personal or professional reputation or harmed their credibility. Defenders perceived online campaigns as an exercise in "character assassination,"[132] and asserted that, "Online hate speech destroys the social identity of people."[133] Defenders report being accosted in restaurants, public bathrooms, and airports as members of the public accuse them of offenses that they read about online, "Many people really believe I am a drug dealer and that I live off reparations for victims for the conflict."[134] One described his lengthy experience with online attacks, "There was a campaign in the press and social media saying 'This lawyer has received Q60,000 ($7,700) and is rich.' It's not true, we took the case pro-bono. But it does damage, people read it in the newspapers and ask, 'Why are you a millionaire?'. . . The hate starts in one place and then spreads. Ordinary people, as well as

[129] Interview with Guatemalan journalist (2019).

[130] Interview with Guatemalan human rights defender (2019).

[131] @LordVaderGT, TWITTER (Dec. 26, 2019) (account now suspended, screenshot with author); @Betocoralg, TWITTER (July 13, 2021), https://twitter.com/Betocoralg/status/1415095182052179973.

[132] Interview with Colombian human rights defender (2020).

[133] Interview with Guatemalan human rights defender (2019).

[134] Interview with Guatemalan human rights defender (2018).

journalists all start to hate you…people start to make comments in bars and restaurants. My daughter went out with friends who said to her, 'Your father is a thief and corrupt lawyer.'"[135] Human rights organizations such as the Colectivo de Abogados José Alvear Restrepo (CAJAR) in Colombia have become so concerned that their public image is tarnished, they have hired a public relations firm to make videos challenging the view that they are "corrupt," and "delinquents" who "buy witnesses."[136]

C. Taking Protective Measures

Over half (54 percent) of defenders reported that attacks on social media have led them to take a variety of protective measures to safeguard their physical security and that of their family. They change their pattern of movements, adopting different routes and modes of travel to and from work. They stop walking with their back to the flow of street traffic. They install bulletproof glass in their car. They stop going out to restaurants and attending public events. A female defender reflected, "Now I always use a pager and travel with someone. I avoid having a fixed schedule. I tell my family, 'I'll call you at 5PM and if I don't, then come and look for me.'"[137] The Inter-American Commission of Human Rights (IACHR) has ordered the Colombian and Guatemalan governments to provide twenty-four-hour police protection to numerous high profile defenders such as Judge Ramón Cadena.[138] During his interview, Cadena expressed his appreciation for the armed policeman standing outside his office but wryly observed, "It diminishes the risk, but nothing will help you if they really want to get you."[139]

D. Interference with Human Rights Work

A majority (54 percent) of defenders indicated that anti-human rights speech online has impeded their effectiveness in their human rights work. Damage to their reputation, they maintain, undermines their credibility as a reliable source, as well as working relationships with clients,

> We represent the victims of the massacre of Mapiripán. They said to the families that we represent that they were deceived

[135] Interview with Guatemalan human rights defender (2019).

[136] Interview with Colombian human rights defender (2020).

[137] Interview with Guatemalan human rights defender (2019).

[138] RESOLUCIÓN 49/2016 MEDIDA CAUTELAR NO. 661-16, COMISIÓN INTERAMERICANA DE DERECHOS HUMANOS (2016). Other defenders receiving IACHR protective measures include Thelma Aldana and Helen Mack of Guatemala and CAJAR and Ricardo Calderón Villegas of Colombia.

[139] Interview with Ramón Cadena, International Commission of Jurists, Guatemala City (2019).

> by us and the Inter-American Court of Human Rights. . . . On social media they accused us of being "white-collar criminals," "thieves" and "corrupt." They say that this person is not a real victim. The victims who we represent have come to doubt our honesty and integrity."[140]

Urban defenders describe encountering hostility when traveling in rural areas, "When our staff go out to communities, they are sometimes met by community members with machetes saying, 'We know you from Facebook. You are traitors receiving funds from international sources.'"[141] In rural Guatemala, villagers have attempted to lynch urban human rights workers based on false stories circulating on social media.[142] Social media attacks can drive defenders out of the human rights sector altogether and thwart future job prospects. After the UN anti-corruption agency (CICIG) ended its work in 2019, there was a coordinated social media campaign threatening former Guatemalan staff of CICG with incarceration and menaced any employer that was considering hiring them, "I had no work after CICIG left and was unemployed. I applied to many jobs including for UN agencies working in Guatemala and was told, 'We don't want anything to do with you guys. Thank you and goodbye.'"[143] At the end of the interview, one human rights attorney threw up his hands and exclaimed, "They made life impossible!"[144]

E. The Connection to Physical Harms

Anti-human rights speech online is seen by 51 percent of interviewees as causally connected to offline attacks. Some defenders made a strong case for a causal nexus, "These social media campaigns block your work, ruin your reputation, and prepare the ground for taking your life,"[145] and, "Hate speech in the United States does not lead to violence. Here [in Colombia], it does, because there is not a gap between what is said and what is done."[146] Defenders advanced many theories about the relationship between anti-human rights speech and offline violence. Some saw a one-to-one causal connection. When asked to identify specific cases in which online speech caused offline harms, Colombian defenders referred to the death threats issued by a shadowy paramilitary group called the "Black Eagles." Their pamphlets which are reposted online customarily announce, "Sentenced To

[140] Interview with Colombian human rights defender (2020).
[141] Interview with Guatemalan human rights defender (2019).
[142] *Id.*
[143] Interview with Guatemalan human rights defender (2020).
[144] Interview with Guatemalan human rights defender (2019).
[145] Interview with Guatemalan human rights defender (2019).
[146] Interview with Colombian human rights defender (2020).

Death. . . . Your time is over. You are going to die."[147] The Black Eagles' death lists have included human rights defenders and organizations who they call "communists," "prostitutes," "thieves," "marijuana smokers," and "people with AIDS" who will be subjected to "social cleansing" (*limpia social*). Guatemalan defenders cite the killing of Jorge Juc Cucul in July 2019 after Facebook posts called him a "robber of energy" and an "enemy of development" because he advocated for the nationalization of the electricity grid.[148] Rural defenders face much higher risks of violence than their urban counterparts, and reports from United Nations agencies have explained that the elevated levels of killings of rural activists correspond with the absence of the rule of law and state authority in rural areas.[149] A Bogotá-based defender stated, "Our organization was named in a leaflet of the Black Eagles, and we went to the police and the prosecutors immediately, but we got no protection from the state. In this case, we didn't have a problem because we are in an urban area, but those who had problems were the rural leaders who have to face the paramilitaries in the territories they control."[150] Guatemalan defenders cited a public speech in May 2018 by President Jimmy Morales that referred to human rights activists as "criminals" and in the next month, eight rural defenders were killed, the highest number in one month that year.[151] Human rights work is especially dangerous for environmental activists opposing hydroelectric dam projects and multinational mining operations. Defenders also believe that private security guards are the most likely initiators of violence, "There is a clear correlation: where there are resources, there is violence."[152]

In contrast, nearly half (49 percent) of interviewees denied that there is a direct causal nexus between online speech and offline harms. These respondents noted that interpersonal violence is generalized in Colombia and Guatemala and that these two countries have among the highest homicide rates in the Americas and the world.[153] Therefore, it is problematic to isolate

---

[147] Águilas Negras, leaflet released in Cauca, Colombia (2016).

[148] Interview with Guatemalan human rights defender (2019); EFE, *Asesinan a un Defensor Indígena y Campesino en Guatemala y Suman 8 Este año*, PRENSA LIBRE (July 26, 2019), https://www.prensalibre.com/guatemala/asesinan-a-un-defensor-indigena-y-campesino-en-guatemala-y-suman-8-este-ano/.

[149] *See* Forst, *supra* note 84, ¶ ¶ 32-36.

[150] Interview with Colombian human rights defender (2020).

[151] Ollantay Itzamná, *¿Quiénes y por qué Están Asesinando a Defensores Comunitarios de Derechos en Guatemala?*, PRENSA COMUNITARIA (June 15, 2018), https://www.prensacomunitaria.org/2018/06/quienes-y-por-que-estan-asesinando-a-defensores-comunitarios-de-derechos-en-guatemala/.

[152] Interview with Guatemalan human rights defender (2019).

[153] In Colombia and Guatemala the homicide rate is 37 and 25 per 100,000, respectively. *See Estimates of Rate of Homicides (per 100,000 Population)*, WORLD HEALTH ORGANIZATION (2021), https://www.who.int/data/gho/data/indicators/indicator-

any causal role for social media, as a prominent Colombian defender who is also a spokesperson for Facebook/Meta, indicated, "Threats do not have physical effects . . . social media companies did not cause the violence. It is endemic in Colombian society."[154] Many defenders nevertheless identified the conditioning effects of anti-human rights rhetoric on social media, "The effects of social media are not usually direct. It is difficult to point to a direct result. The effects are more structural."[155] Another suggested, "These enormous campaigns create the appropriate environment for the attacks."[156] Some defenders maintained that anti-human rights speech did not cause deaths, but justified killings after the fact, "The stigmatizing of defenders naturalizes the killing of their leaders. . . . If someone is killed, people look on Twitter and say, 'Oh they must have done something. They must be FARC [the main Colombian Marxist guerrilla organization].'"[157]

F. The Criminalization of Human Rights Work

About half (49 percent) of defenders perceive a connection between social media and the criminalization of human rights in Colombia and Guatemala. Digital authoritarianism's integration of lawfare and social media is most apparent when a criminal complaint (*denuncia*) is released online in a post.[158] In most cases, however, there is no indictment or arrest warrant. This too can impact defenders by placing them in a legal limbo. While the criminal complaint is outstanding, anti-human rights accounts post that the defender is "under investigation" or a "fugitive from justice," casting a pall over their reputation for probity. A prosecutor investigating corruption of high-level Guatemalan officials described the effects of a social media campaign and criminal complaint against her: "They accused me of money-laundering $15,000 and brought a *denuncia* before the anti-money laundering unit of the Justice Ministry. It was obstruction of justice and a pretext to stop my investigations. The Justice Ministry opened a case and started an investigation, and I had to recuse myself from the case I was working on [. . .] The case against me is still not closed [four years later]. A judge could issue a warrant for my arrest so I am frightened to return to Guatemala. They want to make an example of me."[159] Defenders with children can be especially affected by the legal uncertainty of having a criminal case opened against them,

details/GHO/estimates-of-rates-of-homicides-per-100-000-population.

[154] Interview with Colombian human rights defender (2020).

[155] Interview with Guatemalan human rights defender (2019).

[156] Interview with Guatemalan human rights defender (2019).

[157] Interview with Colombian human rights defender (2020).

[158] Fundación Myrna Mack, *supra* note 99, at 15, 25, 27.

[159] Interview with Guatemalan human rights defender (2020).

> They are using social media to criminalize us . . . to construct us as the enemy and say indigenous people are invaders and violent. . . . It creates a climate of fear. Orders of arrest are not publicly disclosed by law, so we don't know if there is warrant out for our arrest. I fear for my newborn child. If they put me in jail, what will happen to my children?[160]

The same defender also asserted a connection between the criminalization of human rights work and the killing of defenders, "Of the twenty-six defenders killed last year [in Guatemala], all of them had outstanding arrest warrants."[161]

G. Health Effects

Thirty-eight percent of defenders reported a variety of adverse health effects resulting from online harassment. They described insomnia and gastrointestinal problems and they attributed this to the fear and social isolation they have experienced. Most of those suffering from health complications also reported adverse psychological symptoms, and these were most acute for defenders who were under surveillance. One defender, who faced a sustained campaign on social media that accused her of committing crimes and posted pictures of her in public places, pointed to the "psychological effects of being watched, being under constant surveillance."[162] The gender dimensions of online harassment are apparent in the interviews for this study, as women defenders reported generally higher levels of stress after being subjected to online harassment and surveillance.[163] A group of Guatemalan defenders opposing a gold mining operation spoke about the health effects of the coordinated information operation against them,

> There was a campaign of defamation against us. They said our wives were prostituting themselves and finding other men because we were always out protesting. They used the names of the women. They wanted to divide us . . . it confused a lot of people . . . we don't want to all end up sick.[164]

H. Silencing Effects

Twenty-six percent of defenders disclosed that they either ceased to speak

---

[160] Interview with Guatemalan human rights defender (2019).
[161] *Id.*
[162] Interview with Guatemalan human rights defender (2020).
[163] Interview with Colombian human rights defender (2020).
[164] Interview with Guatemalan human rights defender (2019).

in public, publish in the press, moderated their speech or refrained from expressing their views to the fullest. Some have withdrawn from social media or ceased posting either temporarily or permanently, and engaged, in the words of one defender, in "self-censorship."[165] One attorney described a campaign against him on Twitter in which posts revealed his address, disclosed his wife's identity, and conveyed a death threat by posting a picture of a murdered human rights lawyer with the subtitle, "The same will happen to you." When asked if he had altered his activities, he replied, "I should do my work, and what is good for my country. But I need to be careful. I stopped publishing articles and stopped criticizing the government. . . . I went to the US Embassy for a meeting, but I was under surveillance. Within ten minutes there was a picture of me leaving the Embassy on Twitter."[166]

I. Flight from the Country

Eighteen percent of defenders reported either leaving the country temporarily or permanently, moving family members such as children abroad, or making significant plans to leave the country such as arranging travel, taking out a passport and applying for a visa or for asylum in another country, usually the United States. Two former attorneys-general fled Guatemala in part because of a coordinated campaign of harassment and threat on social media. Former attorney-general Claudia Paz y Paz, who successfully prosecuted former President General Ríos Montt for genocide, was forced out of the country and now lives in the United States. Former attorney-general Thelma Aldana, who successfully prosecuted former President Otto Pérez Molina on corruption charges, was granted asylum in the United States in 2020.[167] Leaving the country is a more viable option for urban and professional defenders, and many rural defenders do not possess the resources necessary to apply for asylum. They may still experience physical displacement within the country according to a Colombian defender,

> In the rural areas, it just takes one death threat for people to leave their home and land. This is the product of fear after so many attacks by paramilitaries. They are displaced . . . if they threaten their family, this has the most impact . . . then they might leave the country.[168]

[165] Interview with Colombian human rights defender (2020).

[166] Interview with Guatemalan human rights defender (2019).

[167] *Ex-Guatemala Prosecutor Granted Asylum in U.S.*, AP NEWS (Feb. 24, 2020), https://apnews.com/article/ce4c035ff39ba0f0b2362adaa529194e.

[168] Interview with Colombian human rights defender (2020).

J. Ignoring Online Attacks

Ten percent of defenders indicated that they ignore online attacks and block accounts that harass them. This approach is most prevalent among urban, educated defenders who came to prominence in the era before social media,

> I have been in the struggle for twenty-eight years and I am in my seventies, so I am tough. I block them and don't care what they think. I don't lose sleep; they aren't going to kill me! But young people are affected. Some are more sensitive than others.[169]

Defenders who disregard online threats are also more likely to reject a connection between anti-human rights speech and offline harms, "We have a fluid dialogue with social media companies. The problem of violence is structural and profound. It is not the responsibility of the social networks to change the society. It is their responsibility to educate the population."[170]

VII. THE CONDITIONING EFFECTS OF DIGITAL AUTHORITARIANISM

The central finding from the qualitative interviews conducted for this study is that digital authoritarianism facilitates a range of harmful but usually non-lethal outcomes. Non-lethal harms are seldom included in official government statistics or reports produced by civil society organizations, but they nonetheless have a substantial impact on defenders and their ability to undertake human rights work. Many of the negative effects of anti-human rights speech are inter-locking and reinforce one another; for instance, when reputational damage interferes with human rights work by undermining the credibility of a defender who then engages in self-censorship. These might be termed "indirect effects" of online harassment and threats, but they are as direct as violent, kinetic harms, albeit less visible and measurable. Taken as a whole, they may obstruct human rights work on a daily basis as much as assaults and killings.

Half of defenders interviewed maintain that anti-human rights speech online facilitates the criminalization of their work, underlining the ways in which lawfare and digital authoritarianism are integrated. This is obviously the case when criminal *denuncias* are released in a tweet, but it may be true in a broader sense as well. Defenders made a persuasive case in the interviews

[169] Interview with Guatemalan human rights defender (2019).
[170] Interview with Colombian human rights defender (2020).

for the conditioning effects that shape public discourse. The avalanche of accusation and counteraccusation, false information, and slander, can have a cumulatively corrosive effect on public discourse. Defenders point out that even if they can lodge an effective rebuttal of the accusations against them, online anti-human rights speech erodes the basis for determinations of truth and facticity and thereby destabilize the conditions of knowledge more generally. In the words of one defender, "In the bombardment, no one believes anything, so no one has credibility."[171] The relativization of truth by authoritarian political systems has a long political history.

Even though this study has focused on a wide spectrum of harms, it is important not to dismiss the possibility of a causal nexus between online speech and offline violence against human rights defenders. Although there may be specific cases where a post incites a follower to assault or murder a defender, these instances are rare. Online incitement may create an atmosphere of tolerance of harms against defenders and may enhance the coalitional political identity needed for violence.[172] In this model, anti-human rights speech online conditions a population to oppose human rights defenders and countenance violence against them, rather than directly inciting the public to commit violence themselves, although this is possible too.

Digital authoritarianism has effects that are distinctive from hate speech directed against religious, ethnic, or racial minorities. The Minority Model, which is prevalent in the social science literature and at the United Nations, focuses on incitement of physical violence against a minority population and less on the conditioning effects of speech. The Political Model emphasizes state involvement in censorship and silencing of dissidents but seldom analyzes the macro-level societal effects. This study combines elements of the Political Model and Minority Model to create a theoretical framing that encompasses the full range of impacts of anti-human rights speech online. The primary objective of anti-human rights speech is not to directly incite ordinary citizens to harm or kill defenders, although this does occur. Rather, in the words of a Colombian defender, its aim is to "control the narrative" by stigmatizing defenders and undermining their authority and ability to hold corrupt leaders accountable.[173] Some harms of online speech are related to how it can have a silencing effect on the activities of individual human rights defenders and on public support for their activities. Further, when amplified by platforms, anti-human rights speech creates an environment in which criminalization and violence against defenders are

[171] "En el bombardeo, nadie crea nada, nadie tiene credibilidad." Interview with Guatemalan human rights defender (2019).

[172] Petersen, *supra* note 66.

[173] Interview with Colombian human rights defender (2020).

more likely to be tolerated.[174]

VIII. POLICY RECOMMENDATIONS

Free speech advocates have argued that anti-human rights speech that is merely offensive can be met with long-term strategies of education, digital literacy, and political counter-speech.[175] However, counter-speech presupposes an equality of arms and a scenario in which individual speakers interact in the marketplace of ideas independently of one another and in a political context of rule of law. In many countries of the world, these are not the actual circumstances on the ground. Instead, state security agencies with tremendous institutional capacity are orchestrating mass campaigns of surveillance and harassment against individual journalists, human rights activists, and political dissidents in contexts where there is widespread political and interpersonal violence. Regulation, then, is both lawful and justified when state digital practices break with international laws and norms.

Scholars have accurately pointed out that many of core practices of digital authoritarianism are prohibited under international law. Tamar Megiddo, for instance, draws our attention to the potential violations of an individual's right to privacy and basic democratic freedoms and rule of law by the many digital practices of states.[176] International human rights law is strongly supportive of freedom of expression, but permits limitations that are provided by law,[177] necessary to meet a legitimate objective,[178] and proportionate to the interest to be protected.[179] Article 19(3) of the International Covenant on Civil and Political Rights (ICCPR) states that the only objectives considered legitimate are:

> a. Respect of the rights or reputations of others or
> b. Protection of national security or of public order, or of public health or morals. International human rights law does not permit states to violate the rights of defenders by curtailing their legitimate democratic speech, by damaging their reputation, or by inciting violence against them.

Furthermore, Article 20(2) of the ICCPR explicitly prohibits "any

---

[174] Land & Hamilton, *supra* note 68.

[175] NADINE STROSSEN, HATE: WHY WE SHOULD RESIST IT WITH FREE SPEECH, NOT CENSORSHIP 128 (2018).

[176] Megiddo, *supra* note 6, at 7, 31.

[177] HRC, U.N. DOC. CCPR/C/GC/34 (July 11-29, 2011).

[178] *Id.* ¶ 28, 29.

[179] *Id.* ¶ 34.

advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence." Some forms of state-sponsored anti-human rights speech are prohibited by international law, especially those that draw on national, racial or religious tropes or that incite violence. Instead of making threats and inciting violence against human rights defenders, states have an obligation to take positive steps to protect freedom of expression, equal access to information, and freedom of association for their citizens. The online activities of approximately half of the world's states currently impede the exercise of the basic citizenship rights of their populations.

States must cease and desist from targeting human rights defenders in their information operations and, under the auspices of the United Nations, negotiate, sign, and ratify a Digital Code of Conduct. The UN Digital Code of Conduct should, at a minimum, commit states to ensure transparency in their digital practices, require them to support freedom of expression and the exercise of basic citizenship rights online, and prohibit them from threatening and inciting violence against human rights defenders, and proscribe them from conducting unfettered surveillance and smear campaigns against their citizens.[180]

For their part, technology companies have a duty of care to their over 4 billion users since they presently possess the principal capacity to regulate state-sponsored campaigns of harassment. Social media companies currently detect and disrupt some state information operations, especially when they use networks of bots and engage in "coordinated inauthentic behavior," but they do so in an inconsistent manner. Content posted in Bogotá is not moderated as assiduously as content posted in Boston, and some observers have concluded that platforms have "neglected the rest of the world, fueling hate speech."[181]

The universal scope and application of platforms' content moderation policies mitigates against context-specific policies that can effectively protect human rights defenders. As a Colombian interviewee remarked, "There's a context they don't know . . . they have a universal policy on hate speech for the whole world. But speech is local."[182] Platforms must move away from a one-size-fits-all content moderation policy towards a context-specific approach that is informed by, and responsive to, the circumstances on the ground.[183] Content moderation policies must facilitate a pluralization of

---

[180] *See* Kaye *supra* note 4, ¶ 52-53 (May 28, 2019) (discussing the need for public oversight of surveillance technologies by states).

[181] Cat Zakrzewski, Gerrit De Vynck, Niha Masih & Shibani Mahtani, *How Facebook Neglected the Rest of the World, Fueling Hate Speech and Violence in India*, WASH. POST (Oct. 24, 2021).

[182] Interview with Colombian human rights defender (2020).

[183] Richard Ashby Wilson & Molly K. Land, *Hate Speech on Social Media: Content Moderation in Context*, 52 CONN. L. REV. 1029 (2021); Richard Ashby Wilson & Jordan

speech norms and a decentralization of their operations while conforming to international human rights principles of freedom of expression, transparency, accountability, and respect for due process.

Ideally, this means that human content moderation should be conducted by personnel who are native speakers of the language and who are familiar with the political and cultural context in which the speech is occurring. In at-risk countries where human rights defenders, journalists, and others are repeatedly arrested on criminal charges and killed, social media companies are advised to open local offices and develop strong "trusted partner" relationships with civil society groups, including human rights organizations. Some companies have established internal working groups to monitor developments in countries with a heightened risk of political violence in order to assess how the local situation manifests itself on their platforms, but they need to go further in integrating the offline and online signals into their risk assessment matrix.

Some platforms such as Facebook/Meta and Twitter currently conduct an expedited review of posts flagged by local trusted partners and this practice needs to be adopted more widely. Defenders report that they flag posts that may violate the platforms' terms of service but never received a response. They have sought a line of communication with social media companies, only to be rebuffed. A leading Colombian journalist remarked "We report, and nothing happens. Social media companies are very far from Latin America."[184] In-country offices should be staffed with journalists, human rights attorneys, and political analysts who understand the political context, especially during elections when the risk of public violence is highest. Platforms should not rely on Artificial Intelligence and automated content moderation alone but should integrate external signals identified by the local content and analysis teams into the content moderation matrix. Some companies such as Facebook/Meta include human rights defenders as a protected category and wider adoption of this policy is warranted.[185]

If states succeed in silencing critical voices, undermining anti-

---

Kiper, *Incitement in an Era of Populism: Updating Brandenburg After Charlottesville,*5 U. PENN. J. L. & PUB. AFF. 56 (2020).

[184] Interview with Colombian journalist (2020).

[185] Kari Paul, *Facebook Rule Protects Journalists and Activists as "Involuntary" Public Figures,* THE GUARDIAN (Oct. 13, 2021), https://www.theguardian.com/technology/2021/oct/13/facebook-involuntary-public-figures-journalists-harassment-bullying. *See* Facebook/Meta, *Corporate Human Rights Policy*, ¶ 4 Protecting Human Rights Defenders, https://about.fb.com/wp-content/uploads/2021/04/Facebooks-Corporate-Human-Rights-Policy.pdf (describing its commitment to enacting special protections for human rights defenders) (Accessed 10 April 2022).

corruption campaigns, and "controlling the narrative" on human rights violations through social media, then the prospect for democracy is bleak. As authoritarian and populist politicians make advances worldwide, and as authoritarian states and illiberal democracies take advantage of the affordances of social media to coordinate online information operations against potential sources of opposition, the need to strengthen the protections for civil society actors and independent voices becomes even more pressing.

* *Richard Ashby Wilson* is the Associate Dean of Faculty Development and Intellectual Life, Gladstein Chair of Human Rights, and Board of Trustees Distinguished Professor of Law and Anthropology at the University of Connecticut School of Law.

I appreciate the support I have had for this project and the excellent insights and feedback from colleagues on this article. Specifically, I would like to thank Ginna Anderson, Brittany Benowitz, Jo-Marie Burt, Sandy Coliver, Eric Curwin, Brian Dooley, Juan Franco, Rachel Lopez, Richard Parker, Michael Rubin, Jamie Rowen, Rachel Sieder, and Kimberly Theidon. I am sincerely grateful to the human rights defenders who participated in this study and who, for their protection and security, cannot be named here. Research was supported by the Human Rights Institute and School of Law at the University of Connecticut and by Robert and Mary-Jane Yass. Molly K. Land was a co-researcher in this study and participated in the project design and interviews in Guatemala in 2019. Research assistants Devon Fray, Juan Gutierrez, Emily Karr, Paola Leiva, and Danielle J. Nadeau did stellar empirical work. Adway Wadekar and Danielle J. Nadeau contributed to the statistical analysis. All errors are my own.