Law and Disinformation in the Digital Age

A Project of the Georgetown University Law Center Global Law Scholars

Class of 2023

January 10, 2022

Table of Contents

Introduction

Disinformation has burst onto the world stage as a significant factor in international politics. Although lies, deception, and perfidy are ages-old phenomena, the digital age has facilitated the amplification and manipulation of false information to an unprecedented extent. Disinformation, broadly defined as false information intended to mislead, emanates from both states and non-state actors and affects communities across the globe.

However, acknowledging the importance of disinformation and its consequences invites a host of questions about both disinformation challenges and their potential solutions. How do disinformation campaigns interact with economic, political, social, legal, and technological conditions in targeted communities? How should governments balance regulating information and protecting freedom of expression? What unintended consequences may arise from disinformation laws? What role should private technology and communications firms play in countering disinformation campaigns? How should international institutions address disinformation challenges? How can disinformation exacerbate ethnic tension, armed conflict, and public health crises? How can states use disinformation as a tool of strategic competition? How do bodies of international and domestic law address, preempt, and respond to disinformation?

The diversity of these questions is reflected in the variety of approaches that the authors of this collection take in their essays. This collection of essays, a project of the Georgetown University Law Center's Global Law Scholars class of 2023, examines disinformation from a legal perspective.[1] The essays investigate how individuals, communities, states, and the international community experience and respond to disinformation challenges. While some essays apply

[1] Disclaimer: The contents of individual essays do not necessarily reflect the opinion of Global Law Scholars as a whole or the position of Georgetown University or any of its affiliates.

international legal principles to disinformation, others focus on comparative or foreign law. The collection examines case studies from around the world, using a diversity of sources rarely found in legal scholarship on disinformation.

Together, these essays aim to shed light on the various legal dimensions of international disinformation challenges and contribute to a better understanding of potential solutions to disinformation issues.

Regional Perspectives

Online disinformation has become a global phenomenon, reaching far beyond headline-grabbing events in North America and Europe. However, an excessive focus on disinformation issues in the Global North would obfuscate the nature and gravity of disinformation challenges in the Global South. Disinformation's manifestations and consequences differ based on the diverse political, social, economic, and technological contexts in which it emerges.

This Section takes a regional approach to disinformation, identifying trends based on case studies from Latin America, the Middle East and North Africa, and Sub-Saharan Africa. The included essays examine how disinformation issues interact with regional themes. As the following essays show, analyzing disinformation issues within their regional contexts contributes to a more nuanced understanding of variation in disinformation issues – and potential solutions – around the world. Furthermore, examining regional trends in legal approaches of disinformation helps identify which paradigms may be replicable in other contexts and, perhaps even more importantly, which should be avoided.

State-Sponsored Disinformation: A Look into Latin America

Introduction

Many repressive governments in Latin America have enacted cyber laws, hammered down on freedom of speech, and resorted to other less conventional means to limit people's ability to speak freely and publicly. This, in turn, has opened a void which State governments are able to use to strategically advance their own narratives, usually sowed with disinformation. This is especially true in countries with closed or closing information systems, in which the government controls the flow of the media and information. However, this Chapter focuses instead on how these States use disinformation as a shield by excusing its censorship practices and restriction on freedom of expression to achieve the same end of advancing its own narrative.

Part I of this Chapter establishes the international legal standards that can be used as a defense against State-sponsored disinformation, including international agreements and international organizations. Next, this Chapter analyzes two case studies of repressive state governments in Latin America that have allegedly sponsored and spread disinformation among its people. Part II analyzes the Cuban Government's use of censorship and disinformation to maintain authoritarian control of the country and its people. Part III analyzes the Venezuelan Government's spread of disinformation in the same context. This Chapter, in Part IV, concludes by finding international standards and mechanisms ineffective against State-sponsored use of disinformation.

I. Applicable Legal Standards

This Part looks at two of the existing international standards for monitoring and addressing disinformation: international agreements and international organizations.

a. *International Agreements*

The Universal Declaration of Human Rights' ("UDHR") article 19 provides that, "[e]veryone has the right to freedom of opinion and expression… includ[ing] freedom to hold opinions without interference and to seek, receive and impart information and ideas *through any media and regardless of frontiers*." [1]

While the UDHR is not itself legally binding, much of it has become customary international law.[1] This particular premise has been incorporated into article 19 of the International Covenant on Civil and Political Rights ("ICCPR"). Almost identically to the UDHR, the ICCPR provides that, "[e]veryone *shall* have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice."[2] Furthermore, the cross-border media that article 19 refers to includes all forms of audio-visual, electronic, and internet-based platforms of expression.[3] Venezuela, analyzed in part III, is a party to the ICCPR.[4] Cuba, analyzed in part II, while a signatory, has not ratified the ICCPR and is thus not a party to the treaty.[5] The ICCPR is legally binding and violations can be raised with the International Court of Justice (ICJ).

Article 13 of the American Convention of Human Rights ("ACHR") provides almost identical language, protecting the international right to freedom of thought and expression

[1] Universal Declaration of Human Rights, art. 19, G.A. Res. 217 (III) (Dec. 10, 1948) (emphasis added).
[1] Hurst Hannum, *The Status of the Universal Declaration of Human Rights in National and International Law*, 25 GA. J. INT'L & COMP. L. 287, 289 (1996).
[2] International Covenant for Civil and Political Rights art. 19, Dec. 12, 1966, 999 U.N.T.S. 171 (emphasis added) [hereinafter ICCPR].
[3] UNHRC, *General Comment No. 34: Article 19 (Freedoms of Opinion and Expression)*, 102nd Sess., adopted 12 Sep. 2011, UN Doc CCPR/C/GC/34, https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf.
[4] ICCPR, U.N. Treaty Collection (Sep. 12, 2021), https://treaties.un.org/Pages/ViewDetails.aspx?chapter=4&clang=_en&mtdsg_no=IV-4&src=IND.
[5] *Id.*

regionally.[6] However, the ACHR goes further in that it directs Member States to not approach expression through censoring in advance, but use subsequent imposition of liability instead. It also specifies impermissible restrictions, such as government control of newsprint and other media.[7] However, neither Cuba nor Venezuela are parties to the ACHR.[8]

Both discussed articles within ICCPR and ACHR provide States a limited permissible exception for restricting freedom of expression, including response to concerns of public order and national security.[9] However, at least within the context of the ICCPR, it is a high threshold to overcome, and expression may not be limited unless it is provided by law, and for a legitimate purpose.[10]

b. *International Organizations*

The Organization of American States ("OAS") aims to strengthen peace and security within the Americas and cooperatively promote prosperity and development among Member States.[11]

All thirty-five independent States in the Americas have ratified the OAS Charter and are Member States of the OAS—Venezuela and Cuba included.[12] However, Cuba has a special caveat that "the participation of the Republic of Cuba in the OAS will be the result of a process of dialogue initiated at the request of the Government of Cuba, and in accordance with the practices, purposes,

---

[6] S*ee* Organization of American States, American Convention on Human Rights art. 13, Nov. 22, 1969, O.A.S.T.S. No. 36, 1144 U.N.T.S. 123. Perhaps critically, the key difference between the language in the ACHR article 13 and ICCPR article 19 is that the ICCPR uses "shall"—a term known for conveying legally binding intent.
[7] *Id.*
[8] The Inter-American Court, *ABC of the Inter-American Court of Human Rights*, 1 (2019), https://www.corteidh.or.cr/sitios/libros/todos/docs/ABCCorteIDH_2019_eng.pdf. Venezuela denunciated, which took effect on September 10, 2013.
[9] *See supra*, notes 4, 8.
[10] *See supra*, note 5.
[11] Organization of American States, *Charter of the Organization of American States*, art. 2, 30 April 1948, http://www.oas.org/en/about/purpose.asp.
[12] *See Member States,* ORGANIZATION OF AMERICAN STATES (2021), http://www.oas.org/en/member_states/default.asp.

and principles of the OAS."[13] The reason is that Cuba was banned from voting and participating in the OAS in 1962 despite still being considered a member of the OAS, and though that ban expired already, Cuba has chosen not to accept full readmission into the OAS body.[14]

The OAS provides "guides," which are political commitments.[15] Therefore, though the OAS recommends that Member States honor their political commitments, States are not bound to do so.[16] The recommendations provided by the OAS include guidance on handling disinformation. The OAS has noted how digital and social media act as intermediaries in the flow of information, and while it can ease the flow, it also significantly impacts public debate and "a number of governments—especially authoritarian ones—have tried to pressure them to help shape the flow of information according to their interests."[17] The OAS guide further recommends that:

> "… the States of the region, in line with the standards of the inter-American human rights system, should not establish new criminal types to sanction the dissemination of misinformation or false news. Introducing criminal types, which due to the nature of the phenomenon would be vague or ambiguous, could lead the region back to a logic of criminalizing expressions about officials or people involved in matters of public interest and establishing a tool with a strong chilling effect on the dissemination of ideas, criticism, and information for fear of being subjected to a

[13] *Id.*
[14] *Organization of American States (OAS)***,** NUCLEAR THREAT CENTER, https://www.nti.org/education-center/treaties-and-regimes/organization-american-states-oas/.
[15] Inter-American Judicial Committee, *Guidelines of the Inter-American Juridical Committee on Binding and Non-Binding Agreements*, 123 (2020).
[16] *Id.*
[17] Organization of American States, *Guide to guarantee freedom of expression regarding deliberate disinformation in electoral contexts*, 19 (2019.

criminal process, which would be particularly restrictive in the context of the electoral contest."[18]

Besides disapproving State regulatory frameworks for holding intermediaries responsible for content produced by third parties, other OAS recommendations include promoting universal internet access, protecting journalists and social communicators from violence, and enhancing data privacy and transparency.[19] Therefore, the OAS has many idealistic frameworks for solving State-sponsored disinformation, but, as political commitments, its efforts are not legally binding.

The OAS also houses the Inter-American System of Human Rights, which includes the Inter-American Commission on Human Rights ("IACHR"), Inter-American Court of Human Rights, and Office of the Special Rapporteur for Freedom of Expression ("Special Rapporteur"). Significantly, the role of the Inter-American Court is to interpret and apply the American Convention.[20] The IACHR promotes the protection of human rights in the region and evaluates petitions of violations,[21] while the Special Rapporteur, in addition to crafting reports on the state of freedom of expression in the Americas, advises the IACHR in evaluating freedom of expression petitions.[22] However, Cuba and Venezuela have not accepted the contentious jurisdiction of the Inter-American Court.[23] Similarly, neither Cuba nor Venezuela recognize the ICJ as compulsory.[24]

[18] *Id.* at 20.
[19] *Id.* at 31-36.
[20] *See supra*, note 10 at 4.
[21] *Id.*
[22] *See IACHR Principal Functions,* ORGANIZATION OF AMERICAN STATES (2021), https://www.oas.org/en/iachr/expression/mandate/functions.asp.
[23] *See supra*, note 10 at 4.
[24] *See Declarations recognizing the jurisdiction of the Court as compulsory,* ICJ, https://www.icj-cij.org/en/declarations

II. Cuba Case Study

Cuba features a closing, if not closed, information space where much of the media is controlled by the government. In an already restricted country, the internet is crucial to the expression of Cuban people. As a Human Rights Watch senior researcher noted, "[t]he internet has created a rights revolution in Cuba, allowing people to communicate, report on abuses, and organize protests in ways that were virtually impossible…."[25] However, the Cuban government has encroached on people's access to the internet by enacting a cyber law, and infrastructurally cutting off access to the internet on the island.

This Part will provide the relevant sections of the Cuban Constitution that affect the spread of information, provide background on the current state of affairs in Cuba, and compare how the current situation fairs under applicable domestic and international laws.

a. *Constitutional Limitations*

The Cuban Constitution provides protection that could be used to limit the spread of disinformation—if used correctly. Article 54 of Chapter II of the Cuban Constitution provides that, "[t]he State recognizes, respects, and *guarantees people freedom of thought, conscience, and expression*."[26] Meanwhile, Article 55 provides that, "[t]he State establishes the principles of organization and operation for all means of social communication."[27] However, the Cuban government has, in a way, used the latter article to undermine the former by controlling Cuban people's platforms for communication and expression as will be discussed in the next section.

[25] *Cuba: Telecommunications Decree Curtails Free Speech,* HUMAN RIGHTS WATCH (Aug. 25, 2021), https://www.hrw.org/news/2021/08/25/cuba-telecommunications-decree-curtails-free-speech#.
[26] Cuba Const. ch. II, art. 54.
[27] *Id.* at art. 55.

b. *Current State of Affairs*

In July 2021, widespread anti-government protests erupted in Cuba.[28] The Cuban people grew tired of the authoritarian Government's arbitrary restrictions on human rights, lack of food and medicine, and the Government's response to COVID-19.[29] Though these mass protests were peaceful, "the Cuban Government responded with a brutal strategy of repression designed to instill fear and suppress dissent," including arbitrary arrests.[30] This unprecedented civil unrest in Cuba was reported around the world.[31] However, Cuban president Miguel Diaz-Canel denounced the protests and claimed that, "what the world of seeing of Cuba [was] a lie."[32] It was also reported that bots were widely utilized to quickly spread fake news during the time of the protests.[33]

In response, the Cuban Government has resorted to less conventional means for censoring their people and creating a vacuum for disinformation to spread within. Since the island was given widespread internet access, the government has periodically blocked access to social media including WhatsApp, Instagram, and Facebook.[34] These internet shutdowns have generally occurred during times of turmoil in the country, such as the recent anti-government protest in summer of 2021.[35] These shutdowns exemplify an effort to halt the flow of information within the

---

[28] *Cuba: Peaceful Protesters Systematically Detained, Abused*, HUMAN RIGHTS WATCH (Oct. 19, 2021), https://www.hrw.org/news/2021/10/19/cuba-peaceful-protesters-systematically-detained-abused
[29] *Id.*
[30] *See supra*, note 30.
[31] *See, e.g.,* Kiara Alfonseca, *Why protests in Cuba erupted to historic levels and what protesters want*, ABC NEWS (Jul. 16, 2021), https://abcnews.go.com/International/cuba-protestors-demand-answers-economic-crisis/story?id=78839073; Kaelyn Forde, *Cuba protests: The economic woes driving discontent*, AL JAZEERA, (Jul. 16, 2021), https://www.aljazeera.com/economy/2021/7/16/cuba-protests-the-economic-woes-helping-drive-discontent.
[32] *Cuba president rejects coverage of unrest as a 'lie'*, AL JAZEERA, (JUL. 17, 2021), https://www.aljazeera.com/news/2021/7/17/cuba-president-denounces-unrest-as-a-lie-calls-protest-images.
[33] Khaleda Rahman, *Cuba Becomes Battlefield in Fake News War*, NEWSWEEK (Jul. 22, 2021), https://www.newsweek.com/cuba-becomes-battlefield-fake-news-war-twitter-facebook-1611433.
[34] *Cuban Authorities Block Access to Internet in Response to Protests*, VOA NEWS (Jul. 12, 2021), https://www.voanews.com/a/americas_cuban-authorities-block-access-internet-response-protests/6208178.html (citing Alp Toker, director of Netblocks, a London-based internet monitoring firm and noting that Twitter was not blocked).
[35] *See id.*

country, into the country, and out of the country,[36] thereby limiting the flow of accurate information and creating a vacuum for disinformation to thrive.

Further in its response, the Cuban Government, in August 2021, passed a telecommunications decree, titled "Decree-Law 35," which limits online content.[37] It does so by placing a standard on digital users of whether their content is offensive, "upset[s] public order,"[38] or impacts "collective security," "general well-being," or "public morality."[39] Individuals who try to "subvert the constitutional order" may be deemed cyberterrorists.[40] The media has termed this move as "digital repression," calling it "Orwellian" in nature.[41] Yet, the Cuban Government justifies its action by claiming that the purpose of the decree is to counteract "misinformation and cyber lies."[42]

c. *Cuba's Response Analyzed Under Domestic and International Law*

The Decree-Law 35 seemingly violates Article 54 of the Cuban Constitution, as it can be used as a means for limiting people's freedom of thought or expression under the guise of disrupting public order or collective security.[43] However, it is consistent with Article 55 of the Constitution in that the media is guided by the principles established by the State and the Cuban Government imposed this cyber principle.[44] While the constitutionality of the decree is ambiguous, the decree directly contradicts the recommendations of the OAS.

---

[36] *See id.*
[37] *See Cyber Law Gives Cuba New Way to Silence Critics, Analysts Say*, VOA NEWS (Aug. 12, 2021), https://www.voanews.com/a/press-freedom_cyber-law-gives-cuba-new-way-silence-critics-analysts-say/6210050.html.
[38] *Id.*
[39] *See supra*, note 27.
[40] *See supra*, note 39.
[41] *Id.*
[42] *Id.*
[43] *See supra*, note 28.
[44] *See supra*, note 29.

In contrast to the OAS's recommendation to not criminalize misinformation,[45] the Cuban government is doing exactly that. The decree is imposing punishment on intermediaries (the Cuban people), which can be used to target opposition and exacerbate corrupt practices and human rights violations.[46]

More broadly, the Decree-Law 35, as well as the intentional internet outages, violate Article 19 of the ICCPR and Article 13 of the ACHR, because the Cuban Government uses these methods to curtail free expression.[47] Yet, these restrictions on the right are unlikely "necessary and proportionate to achieve a legitimate goal, such as the protection of national security or the rights of others."[48] A court, such as the Inter-American Court or ICJ, would have to have to use a proportionality analysis to determine whether Cuba is legally permitted to restrict expression the way it is doing so. However, Cuba is not a party to the Inter-American Court nor the ICJ, and it is unknown if or when such a case would arise.

III. Venezuela Case Study

Venezuela is another example of a closing information environment.[49] The Venezuelan Government controls much of the media and has sought to restrict the people's access to information.[50] Moreover, the Venezuelan government has used the spread of disinformation and misinformation as a vindication for limiting free speech.

---

[45] *See supra*, note 19.
[46] *See id.*
[47] *See supra*, note 27.
[48] *Id.*
[49] Margarita Seminario, *Free Press, Fake News, and Repression during Covid-19: Venezuela and Nicaragua*, CSIS (Jun. 4, 2020), https://www.csis.org/analysis/free-press-fake-news-and-repression-during-covid-19-venezuela-and-nicaragua.
[50] *See Venezuela, Freedom on the Net*, FREEDOM HOUSE (2020), *https://freedomhouse.org/country/venezuela/freedom-net/2020*

This Part will first provide the relevant sections of the Venezuelan Constitution that affect the spread of information, provide background on the current state of affairs in Venezuela, and compare how the current situation fairs under applicable domestic and international laws.

a. *Constitutional Limitations*

Venezuela's Constitution, Chapter III, Article 57, provides that "[e]veryone has the right to express freely his or her thoughts, ideas or opinions orally, in writing *or by any other form of expression*, and to use for such purpose any means of communication and diffusion, and no censorship shall be established . . .."[51] Venezuela's Constitution, Chapter III, Article 58 additionally reads that, "[c]ommunications are free and plural, and involve the duties and responsibilities indicated by law. Everyone has the right to timely, truthful and impartial information, without censorship, in accordance with the principles of this Constitution . . .."[52]

Thus, the Constitution provides protection for freedom of expression and speech, but the current state of affairs demonstrates a reluctance to comply with it.

b. *Current State of Affairs*

Since 2014, Venezuela has been spiraling into an economic collapse, creating a human rights crisis.[53] There have been widespread power outages, scarcity of food and medicine, and anti-government protests, as also seen in Cuba.[54] In Venezuela, television is mostly State-run, and the Government banned the few independent news sources (television and radio) from reporting on the Venezuelan crisis.[55] There has also been a clamp down on internet freedom, which has led the

[51] Venez. Const. ch. III, art. 5.
[52] *Id.* at art. 58.
[53] Ciara Nugent, *'Venezuelans Are Starving for Information.' The Battle to Get News in a Country in Chaos*, TIME (Apr. 16, 2019), https://time.com/5571504/venezuela-internet-press-freedom/
[54] *See id.* (reporting about Cuba from the United States of America); *supra*, note 30.
[55] *See supra*, note 52.

country to be labeled "not free."[56] This follows a history of "Chavista governments dismantl[ing] the country's independent media, stifl[ing] key opposition outlets, and us[ing] public funds to disseminate state propaganda."[57] For example, in 2010, the Chávez regime passed laws that expanded state-control of the media, and empowered the government to punish critics.[58]

In response to the ongoing crisis, the Venezuelan Government has continued to undermine freedom of information and spread disinformation. The pandemic has exposed how the current Venezuelan government (the Maduro regime) has used disinformation as both a weapon and a shield. The government has arrested and/or attacked journalists and other press workers who have reported on the virus or criticized the Government's response.[59] However, the Venezuelan government has excused these arrests by claiming that the arrestees were spreading dangerous disinformation surrounding the virus.[60] This justification is similar to that used by the Cuban Government to enact its cyber decree.[61]

c. *Venezuela's Response Analyzed Under Domestic and International Law*

Unlike the constitutional ambiguity seen in the Cuba case, Venezuela's crackdown on freedom of speech and expressions appears to plainly violate its Constitutional guarantees. Yet, Venezuela's actions may align with the restrictive laws passed by Chávez in 2010.

Further, much like Cuba's cyber decree providing punishments inconsistent with the OAS recommendations, Venezuela's arrests in response to alleged disinformation are also inconsistent with the OAS guidance. While the OAS counsels governments against criminalizing

[56] *See supra*, note 51.
[57] *Id.*
[58] *Venezuela: Legislative Assault on Free Speech, Civil Society*, HUMAN RIGHTS WATCH (Dec. 22, 2010), https://www.hrw.org/news/2010/12/22/venezuela-legislative-assault-free-speech-civil-society.
[59] *See id.*
[60] *See id.*
[61] *See supra*, note 39.

disinformation, the government is arresting individuals for allegedly spreading disinformation. It again can be used to strategically target opposition and silence people.

More broadly, it is also inconsistent with article 19 of the ICCPR and article 13 of the ACHR, as seen in the Cuba case study. Yet, despite the actions of both the Cuban and Venezuelan governments being inconsistent with these international standards, the countries are not parties to the ACHR, and only Venezuela is a party to the ICCPR. Further, neither of the countries grant automatic jurisdiction to the courts that adjudicate these standards—the Inter-American Court and ICJ respectively.

IV. Conclusion

The actions of the Cuban and Venezuelan governments demonstrate a correlation between authoritarianism, censorship, and spread of disinformation. In closed and closing free speech areas, like Cuba and Venezuela, the spread of accurate information becomes more difficult. To reduce the impact of disinformation, therefore, it is important to keep those environments open as is desired by international laws governing freedom of expression.

Governments using disinformation as a shield is particularly dangerous to freedom of expression and access to information. If, like Cuba and Venezuela, a country's government uses false news and disinformation as a vindication for censoring people, it can disguise its true efforts of silencing opposition and other truths that it disagrees with. For this reason, the OAS recommendation that countries should not impose new punishments or sanctions for false news and information is crucial. However, when the government already controls what "accurate" news is, and the OAS recommendations lack legal enforcement, it is unlikely to influence the actions of State governments.

Additionally, the current international mechanisms are not sophisticated enough to control State-sponsored disinformation and repression that allows for disinformation given the consent-based nature of international law. As noted throughout this Chapter, Cuba and Venezuela are not parties to the ACHR, only Venezuela ratified the ICCPR, and neither have consented to automatic jurisdiction of the Inter-American Court or ICJ.

Combined, this means that State governments like Cuba and Venezuela choose what international standards on disinformation and freedom of expression to comply with or ignore and choose whether to consent to adjudication or enforcement by a court. In conclusion, it grants governments the autonomy to use disinformation if they so choose to.

Digital Authoritarianism and Disinformation Laws in the Middle East and North Africa

Striking a balance between the freedom of expression and legitimate government interests in limiting speech lies at the heart of regulating disinformation.[1] In the Middle East and North Africa (MENA), home to some of the world's most restrictive States in terms of freedom of expression and civil liberties,[2] disinformation laws have facilitated government repression of journalists, human rights defenders, and opposition activists. With COVID-19 escalating disinformation concerns, many MENA States have enacted new disinformation measures and increased enforcement of existing laws against people who allegedly spread disinformation or misinformation about the pandemic.

This Chapter examines MENA disinformation measures through the lenses of digital authoritarianism[3] and authoritarian adaptation.[4] The Chapter argues that authoritarian governments in the region wield counter-disinformation rationales in a manner that restricts freedom of expression, suppresses dissent, and targets legitimate speech. MENA disinformation laws, the enactment and enforcement of which has accelerated in recent years, often contain broadly defined offenses that allow governments to act as arbiters of the truth.

---

[1] *See generally* David Goldberg, *Responding to "Fake News": Is There an Alternative to Law and Regulation?*, 47 Sw. L. Rev. 417 (2018); Andrei Richter, *Fake News and Freedom of the Media*, 8 J. Int'l Media & Ent. L. 1 (2019); Fernando Nuñez, *Disinformation Legislation and Freedom of Expression*, 10 UC Irvine L. Rev. 783 (2020).
[2] Sarah Repucci & Amy Slipowitz, *Freedom in the World 2021: Democracy Under Siege*, Freedom House (2021), https://freedomhouse.org/report/freedom-world/2021/democracy-under-siege.
[3] *See generally* Tiberiu Dragu & Yonatan Lupu, *Digital Authoritarianism and the Future of Human Rights*, Int'l Org., Accepted (2021).
[4] Marc Lynch, *Digital Activism and Authoritarian Adaptation in the Middle East*, Project on Middle E. Pol. Sci. (August 2021), https://pomeps.org/wp-content/uploads/2021/08/POMEPS_Studies_43_Draft3-1.pdf.

Part I examines international and regional human rights standards relevant to disinformation. Part II traces the rise of digital authoritarianism in the Middle East and North Africa after the Arab uprisings of 2011. Part III analyzes the enactment of anti-disinformation measures first as global fake news narratives grew and then during the COVID-19 pandemic. Part III then discusses the enforcement of disinformation measures and spotlights some cases in which civil society opposition blocked the passage of disinformation laws. Finally, Part IV summarizes the Chapter's findings and offers lessons learned.

I. International and Regional Law on Disinformation

Human rights law offers a framework to consider MENA disinformation measures. Well-established freedom of expression norms invite scrutiny of measures that excessively restrict freedom of expression.

The International Covenant on Civil and Political Rights (ICCPR) contains two articles regarding freedom of expression.[5] Article 19 establishes that everyone shall have the right to hold opinions without interference and the right to freedom of expression.[6] Except for some States in the Persian Gulf, most MENA countries have ratified the ICCPR.[7] The ICCPR includes both permissible and mandatory restrictions on the right to freedom of expression. Section 3 of Article 19 recognizes two valid restrictions on the right to the freedom of expression, allowing laws that necessarily restrict expression for 1) the respect of the rights or reputations of others; or 2) the protection of national security, public order, or public health or morals.[8] Article 20 offers two

[5] International Covenant on Civil and Political Rights (ICCPR) art. 19, Dec. 16, 1966, S. Exec. Rep. 102-23, 999 U.N.T.S. 171.
[6] *Id*.
[7] *Id.*
[8] *Id.*

prohibitions on specific types of expression: propaganda for war and "advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence."[9]

However, the Human Rights Committee emphasized in General Comment No. 34 on Article 19 that "restrictions must not be overbroad" and must adhere to the principles of necessity and proportionality.[10] The General Comment further observed that "[t]he principle of proportionality has to be respected not only in the law that frames the restrictions but also by the administrative and judicial authorities in applying the law."[11] The Human Rights Committee emphasized "penalization of a media outlet, publishers or journalist[s] solely for being critical of the government… can never be considered… a necessary restriction of freedom of expression."[12]

Various sources of soft law expand on the ICCPR's freedom of expression provisions. The UN Human Rights Council adopted Resolution 20/8 regarding the promotion, protection, and enjoyment of human rights on the internet in 2012,[13] affirming "the same rights that people have offline must also be protected online, in particular freedom of expression."[14] In Resolution 26/13 of 2014, the United Nations General Assembly took up the same topic, affirming the Human Rights Council's finding regarding freedom of expression online.[15] In 2016, the Human Rights Council returned again to the issue of human rights on the internet with Resolution 32/13.[16] The resolution reaffirmed the findings of 20/8 and 26/13 and condemned "measures to intentionally prevent or

---

[9] *Id.* at art. 20.
[10] UNHRC, General Comment No. 34: Article 19 (Freedoms of opinion and expression), 102th Sess, adopted 12 Sept 2011, UN Doc CCPR/C/GC/34, https://tbinternet.ohchr.org/_layouts/15/treatybodyexternal/Download.aspx?symbolno=CCPR%2fC%2fGC%2f34&Lang=en. The Human Rights Committee defined proportionality to mean that restrictive measures must be appropriate to achieve their protective function, the least intrusive instrument means to achieve their protective function, and proportionate to the interest to be protected.
[11] *Id*.
[12] *Id*.
[13] U.N.H.R.C. Res. 20/8, U.N. Doc. A/HRC/RES/20/8 (Jul. 16, 2012).
[14] *Id*.
[15] U.N.G.A. Res. 26/13, U.N. Doc. A/HRC/RES/26/13 (Jul. 14, 2014).
[16] U.N.H.R.C. Res. 32/13,, U.N. Doc. A/HRC/RES/32/13 (Jul. 1, 2016).

disrupt access to or dissemination of information online in violation of international human rights law" and abuses "committed against persons for exercising their human rights and fundamental freedoms on the Internet."[17]

In 2017, the United Nations Office of the High Commissioner for Human Rights, alongside three regional bodies, released a joint declaration regarding freedom of expression and disinformation. The declaration encouraged the abolition of "[g]eneral prohibitions on the dissemination of information based on vague and ambiguous ideas, including 'false news' or 'non-objective information'" labeling them "incompatible with international standards for restrictions on freedom of expression."[18]

The regional human rights treaties in the Middle East and North Africa also include provisions related to freedom of expression. The Arab Charter on Human Rights, ratified by thirteen countries, "ensure[s] the right to information, freedom of opinion and freedom of expression, freedom to seek, receive and impart information by all means, regardless of frontiers."[19] Mirroring the ICCPR, the Charter recognizes limitations on the freedom of expression but stipulates that it "shall only be subjected to restrictions necessary for the respect of the rights or reputation of others and for the protection of national security or of public order, health or morals."[20] The African Charter on Human and Peoples' Rights, ratified by six MENA States,

---

[17] *Id*.
[18] U.N. Special Rapporteur on Freedom of Opinion and Expression et. al., Joint Declaration on Freedom of Expression and "Fake News," Disinformation and Propaganda, U.N. Doc. FOM.GAL/3/17 (Mar. 3, 2017), https://www.osce.org/fom/302796?download=true.
[19] League of Arab States, Arab Charter on Human Rights, May 22, 2004, reprinted in 12 Int'l Hum. Rts. Rep. 893 (2005) (entered into force Mar. 15, 2008) [hereinafter Arab Charter]. For an English translation of the Arab Charter, see the University of Minnesota Human Rights Library, available at http://www1.umn.edu/humanrts/instree/loas2005.html.
[20] *Id*.

affirms that "Every individual shall have the right to express and disseminate his opinions within the law."[21]

II. The Development of Digital Authoritarianism After the Arab Uprisings

Digital media and technology played a central role in the Arab uprisings of 2011 and the authoritarian consolidation that followed. After the Tunisian fruit-seller Mohammed Bouazizi died by self-immolation in the town of Sidi Bouzid, videos of the event spread around the country, sparking the Jasmine Revolution. In the following days and weeks, Al Jazeera footage of protests spurred on uprisings throughout the region. Social media platforms including Facebook and Twitter served as key tools in organizing protests.[22] As protests gained momentum, so did narratives of "Facebook revolutions" and claims that the internet would democratize the Middle East and North Africa.[23]

However, analysts have since drawn attention to the rise of digital authoritarianism in the region.[24] While popular movements can use technology to facilitate political organizing, regimes also use technology for repression. For example, new digital surveillance technology allows

[21] Organization of African Unity, African [Banjul] Charter on Human and Peoples' Rights art. 9, June 27, 1981, 21 I.L.M. 58 (1982).
[22] *See generally* Kara Alaimo, *How the Facebook Arabic Page "We Are All Khaled Said" Helped Promote the Egyptian Revolution*, July-December 2015 Social Media + Society at 1 (2015); *see also* Heather Brown, Emily Guskin & Amy Mitchell, *The Role of Social Media in the Arab Uprisings*, Pew Research Center, Nov. 28, 2012, https://www.pewresearch.org/journalism/2012/11/28/role-social-media-arab-uprisings/.
[23] *See, e.g.*, Jose Antonio Vargas, *Spring Awakening*, N.Y. Times (Feb. 17, 2012), https://www.nytimes.com/2012/02/19/books/review/how-an-egyptian-revolution-began-on-facebook.html.
[24] *See, e.g.*, Afef Abrougui, *Digital Authoritarianism in the GCC and its Broader Regional Consequences*, Carnegie Endowment for International Peace (Oct. 19, 2021), https://carnegieendowment.org/2021/10/19/digital-authoritarianism-in-gcc-and-its-broader-regional-consequences-pub-85511; *POMEPS Studies 43: Digital Activism and Authoritarian Adaptation in the Middle East*, Jadaliyya (Aug. 12, 2021), https://www.jadaliyya.com/Details/43203/POMEPS-Studies-43-Digital-Activism-and-Authoritarian-Adaptation-in-the-Middle-East.

authorities to surveil critics,[25] governments can censor online content,[26] and some States have mastered the use of digital platforms to disseminate propaganda.[27]

Online speech has emerged as a principal arena for political contestation, forming a virtual Arab street.[28] In a process of authoritarian adaptation, regimes developed new tools to muffle critical voices. Governments have long used allegations of libel, terrorism, causing strife, disrespecting the State, disturbing relations with a foreign State, damaging national unity, and the like to silence critics. As conversations moved online, legal mechanisms followed, with cybercrime laws often taking aim at online speech. The Arab uprisings prompted early anti-disinformation laws as States honed tactics of digital authoritarianism. For example, Oman's sultan issued a cybercrime decree in 2011,[29] and the U.A.E. passed its cybercrime law in 2012.[30]

As Jamal Khashoggi wrote in his last publication before his murder, "There was a time when journalists believed the Internet would liberate information from the censorship and control associated with print media. But these governments, whose very existence relies on the control of

[25] *See* Jon Hoffman, *Espionage and repression in the Middle East courtesy of the West*, Open Democracy (May 15, 2020), https://www.opendemocracy.net/en/north-africa-west-asia/espionage-and-repression-middle-east-courtesy-west/; *see, e.g., Pegasus Project: End export of surveillance technology to MENA autocratic governments*, MENA Rts. Group (Jul. 26, 2021), https://menarights.org/en/articles/pegasus-project-end-export-surveillance-technology-mena-autocratic-government.

[26] *See, e.g.*, Elissa Miller, *Egypt leads the pack in internet censorship across the Middle East*, Atlantic Council (Aug. 28, 2018), https://www.atlanticcouncil.org/blogs/menasource/egypt-leads-the-pack-in-internet-censorship-across-the-middle-east/; Renée DiResta, Josh A. Goldstein & Shelby Grossman, *Middle East Influence Operations: Observations Across Social Media Takedowns* , Project on Middle E. Pol. Sci. (August 2021), https://pomeps.org/wp-content/uploads/2021/08/POMEPS_Studies_43_Draft3-1.pdf.

[27] DiResta, Goldstein & Grossman, *supra* note 26, at 91.

[28] Mohamed Zayani, *Social Movements in the Digital Age: Change and Stasis in the Middle East,* Eur. Inst. of the Mediterranean (2019), https://www.iemed.org/publication/social-movements-in-the-digital-age-change-and-stasis-in-the-middle-east/.

[29] Royal Decree No. 12 of 2011 (Issuing the Cybercrime Law), *al-Jarīdah al-Rasmīyah*, no. 929, 6 February 2011. An English translation is available here: https://www.qcert.org/sites/default/files/public/documents/om-ecrime-issuing_the_cyber_crime_law-eng-2011.pdf.

[30] Bassam Za'za', *Understanding UAE's cybercrime law and penalties*, Gulf News (Sep. 12, 2015), https://gulfnews.com/going-out/society/understanding-uaes-cybercrime-law-and-penalties-1.1564565.

information, have aggressively blocked the Internet."[31] Disinformation laws have emerged as a new method for digital authoritarianism after the Arab uprisings, with fake news narratives and the COVID-19 pandemic facilitating their adoption across the region.

III. Use of Disinformation Laws for Repression

The last decade has seen two trends accelerate disinformation legislation in the region. First, a normalization of narratives about fake news and concerns about the political consequences of disinformation since 2016 facilitated the passage of laws targeting individuals found to spread false information. Second, concerns about reliable public health information during the COVID-19 pandemic have resulted in new counter-disinformation measures and enhanced enforcement of existing laws. In both cases, States have wielded disinformation laws against journalists, human rights defenders, and political opponents.

A. The Rise of Disinformation Legislation, 2017-2020

During the administration of U.S. president Donald Trump, narratives accusing journalists of "fake news" gained traction on the global stage.[32] According to a 2020 report from the Committee for the Protection of Journalists, at least twenty-six countries enacted fake news laws from January 2017 to May 2019, which the Committee partially attributed to the Trump administration's perceived antagonism towards journalists.[33] Meanwhile, legitimate concerns

[31] Jamal Khashoggi, *What the Arab world needs most is free expression*, Wash. Post (Oct. 17, 2018), https://www.washingtonpost.com/opinions/global-opinions/jamal-khashoggi-what-the-arab-world-needs-most-is-free-expression/2018/10/17/adfc8c44-d21d-11e8-8c22-fa2ef74bd6d6_story.html.
[32] *See* Uri Friedman, *The Real-World Consequences of 'Fake News,'* The Atlantic (Dec. 23, 2017), https://www.theatlantic.com/international/archive/2017/12/trump-world-fake-news/548888/; *The global reach of Trump's 'fake news' outrage*, Wash. Post (Nov. 19, 2019), https://www.washingtonpost.com/opinions/global-opinions/trump-is-spreading-his-fake-news-rhetoric-around-the-world-thats-dangerous/2019/11/19/a7b0a4c6-0af5-11ea-97ac-a7ccc8dd1ebc_story.html.
[33] *The Trump Administration and the Media*, Comm. to Protect Journalists (Apr. 16, 2020), https://cpj.org/reports/2020/04/trump-media-attacks-credibility-leaks/#9.

about the political consequences of disinformation mounted after Russian interference in the 2016 U.S. election.[34]

Many MENA leaders took President Trump's rhetoric as a sign that clamping down on fake news would be an acceptable justification for restricting speech. Egypt, Saudi Arabia, and Turkey ranked among the worst jailers of journalists in 2019.[35] In the context of eroding international human rights norms, States adopted fake news and disinformation laws limiting journalistic freedom.

Egypt and Qatar, among other States, passed disinformation legislation as fake news narratives gained momentum. Egypt's Parliament passed Law No. 175 of 2018 on Anti-Cybercrimes and Information Technology Crimes,[36] which treats individuals with over 5,000 social media followers as media outlets and subjects them to prosecution for publishing fake news.[37] The law authorized the Supreme Council for Media Regulation to sanction journalists or publications that spread fake news.[38] On January 8, 2020, Qatar amended the Penal Code with a criminal disinformation statute.[39] The provision, contained in Article 136 of the Code, criminalized the broadcasting, publishing or republishing "false or biased rumours, statements or news, or inflammatory propaganda, domestically or abroad, with the intent to harm national interests, stir up public opinion, or infringe on the social system or the public system of the State."[40] The law authorized punishments of up to five years in prison or a fine equivalent to $27,473.[41] Both the

[34] Nina Jankowicz, *How to Defeat Disinformation*, Foreign Affs. (Nov. 19, 2020), https://www.foreignaffairs.com/articles/united-states/2020-11-19/how-defeat-disinformation.
[35] *The Trump Administration and the Media*, *supra* note 33.
[36] Law No. 175 of 2018 (Countering Cybercrimes), *al-Jarīdah al-Rasmīyah*, vol. 32, 14 August 2018.
[37] *Id*.
[38] Fatma Khaled, *Egypt Begins Legal Crackdown on "Fake News"*, Wash. Rep. on Middle E. Aff., Mar. 2019, at 30.
[39] *Qatar: Repressive new law further curbs freedom of expression*, Amnesty Int'l (Jan. 20, 2020), https://www.amnesty.org/en/latest/news/2020/01/qatar-repressive-new-law-further-curbs-freedom-of-expression/.
[40] *Id*.
[41] *Qatar changes penal code to include 'false news' law*, Comm. to Protect Journalists (Jan. 21, 2020), https://cpj.org/2020/01/qatar-penal-code-false-news/.

Egyptian and the Qatari laws contained overbroad definitions of fake news and vague intent requirements, allowing for inconsistent enforcement.

The rising salience of fake news narratives globally during the Trump administration ushered in a new round of legislation. Now, as the COVID-19 pandemic rages on, a third period of disinformation has become discernible, with states adopting, or attempting to adopt, legislation to ostensibly curb COVID-19-related disinformation.

B. The Escalation in Disinformation Legislation During COVID-19 (2020-2021)

The COVID-19 pandemic has exacerbated concerns about disinformation. As World Health Organization director Tedros Adhanom Ghebreyesus said in February 2020, "But we're not just fighting an epidemic; we're fighting an infodemic. Fake news spreads faster and more easily than this virus, and is just as dangerous."[42]

Governments and private entities began restricting misinformation and disinformation about COVID-19 in spring 2020. Major social media platforms deployed fact-checking measures[43] and implemented content take-down policies.[44] Around the world, governments passed laws regulating fake news and disinformation.[45]

---

[42] Tedros Adhanom Ghebreyesus, *Munich Security Conference*, World Health Organization (Feb. 15, 2020), https://www.who.int/director-general/speeches/detail/munich-security-conference.

[43] Keren Goldshlager and Orlando Watson, *Launching a $1M Grant Program to Support Fact-Checkers Amid COVID-19*, Facebook (Apr. 30, 2020), https://www.facebook.com/journalismproject/coronavirus-grants-fact-checking.

[44] Rachel Lerman, *Facebook says it has taken down 7 million posts for spreading coronavirus misinformation*, Wash. Post (Aug. 11, 2020), https://www.washingtonpost.com/technology/2020/08/11/facebook-covid-misinformation-takedowns/.

[45] Jamie Wiseman, *Rush to pass 'fake news' laws during Covid-19 intensifying global media freedom challenges*, Int'l Press Inst. (Oct. 22, 2020), https://ipi.media/rush-to-pass-fake-news-laws-during-covid-19-intensifying-global-media-freedom-challenges/; Sara Fischer, *"Fake news" laws on the rise globally during the coronavirus pandemic*, Axios (May 26, 2020), https://www.axios.com/fake-news-laws-coronavirus-pandemic-4def8720-9ad8-4b8d-abfa-762581865463.html.

The MENA has faced genuine disinformation challenges during the COVID-19 pandemic. Already prone to conspiracy theories,[46] the region saw a significant uptick in public health-related disinformation on social media. In Egypt, for instance, false claims about the novel coronavirus, treatments, government responses, and conspiracy theories have circulated online, including Russian-backed disinformation.[47] Vaccine disinformation in Arabic proliferated on Facebook, YouTube, and Twitter across the MENA region.[48]

In practice, many governments have taken the COVID-19 pandemic as an opportunity to restrict expression.[49] States' COVID-19 crackdowns were not limited to the digital space. Algeria and Sudan restricted outdoor gatherings, curbing ongoing protest movements.[50] MENA States, including Jordan, Morocco, Oman, and Yemen, have prohibited the distribution of newspapers during the pandemic.[51]

MENA governments have accelerated the passage and enforcement of disinformation statutes during the pandemic. As disinformation concerns spread with COVID-19, an international receptiveness to countering disinformation provided an opening for authoritarian governments in the Middle East and North Africa to further restrict freedom of expression.

---

[46] *See generally* Matthew Gray, *Explaining Conspiracy Theories in Modern Arab Middle Eastern Political Discourse: Some Problems and Limitations of the Literature*, 17 Critique: Critical Middle E. Stud. 155 (2008).
[47] Joey Shea, *Misinfo, Disinfo, and Fake News in Egypt's COVID-19 "Infodemic",* Tahrir Inst. for Middle E. Pol'y (May 8, 2020), https://timep.org/commentary/analysis/misinfo-disinfo-and-fake-news-in-egypts-covid-19-infodemic/.
[48] Ciarán O'Connor & Moustafa Ayad, *MENA Monitor: Arabic COVID-19 Vaccine Misinformation Online*, Inst. for Strategic Dialogue (Apr. 2021), https://www.isdglobal.org/wp-content/uploads/2021/04/MENA-Covid-Vaccine-Misinformation-Monitor-1.pdf.
[49] Charles W. Dunne, *Authoritarianism and the Middle East in the Time of COVID-19*, Arab Center Washington DC (Apr. 2, 2020), https://arabcenterdc.org/resource/authoritarianism-and-the-middle-east-in-the-time-of-covid-19/; Alexis Thiry, *Will COVID-19 Create a Human Rights Crisis in the Middle East and North Africa?,* Just Sec. (Oct. 1, 2020), https://www.justsecurity.org/72643/will-covid-19-create-a-human-rights-crisis-in-the-middle-east-and-north-africa/.
[50] *Virus brings halt to more than year of weekly Algeria rallies*, France 24 (Mar. 20, 2020), https://www.france24.com/en/20200320-virus-brings-halt-to-more-than-year-of-weekly-algeria-rallies.
[51] *COVID-19 Civic Freedom Tracker*, Int'l Ctr. for Not-For-Profit Law, https://www.icnl.org/covid19tracker/?issue=9 (last accessed Nov. 3, 2021).

In Algeria, the Hirak protest movement began in February 2019 after President Abdelaziz Bouteflika, then eighty-two years old, announced his candidacy for another term as Algeria's president.[52] Since then, the Algerian government has responded with repression, including arrests on position figures and bans on protests.[53] Algeria's restrictions have included new disinformation measures. On April 22, 2020, Algeria's Parliament amended the Penal Code with Law 20-06 to add an article punishing anyone who voluntarily spreads "false or slanderous" information "likely to harm state security or public order" with up to three years in prison and fines equivalent to around $2,322.[54] The broad construction of "likely to harm" national security leaves interpretation to the Algerian State, which has interpreted political opposition during the Hirak as threatening to national security. Without a more precise definition of "false or slanderous," the government serves as the ultimate arbiter of truthful information. In November 2020, the Algerian government issued a decree tightening state control over online media.[55] Minister of Communications Ammar Belhimer cited combating the spread of "rumours, fake news and fake videos" to justify the new measures.[56]

On June 24, 2020, Mauritania's National Assembly passed the Law on Combating Manipulation of Information, which addresses the publication of "inaccurate information" or "fake news" on social media.[57] The law criminalizes "spreading false information with the goal of

---

[52] Andrew G. Farrand, *Two years on, Algeria's Hirak is poised for a rebirth*, Atlantic Council (Feb. 16, 2021), https://www.atlanticcouncil.org/blogs/menasource/two-years-on-algerias-hirak-is-poised-for-a-rebirth/.
[53] Ilhem Rachidi, *Helpless Hirak? Democratic Disappointments in Algeria*, Carnegie Endowment for Int'l Peace (June 10, 2021), https://carnegieendowment.org/sada/84739.
[54] Yasmine Kacha, *In a post-COVID-19 world, "fake news" laws, a new blow to freedom of expression in Algeria and Morocco/Western Sahara?*, Amnesty Int'l (May 29, 2020), https://www.amnesty.org/en/latest/news/2020/05/in-a-post-covid19-world-fake-news-laws-a-new-blow-to-freedom-of-expression-in-algeria-and-morocco-western-sahara/; Law 20-06, *al-Jarīdah al-Rasmīyah*, no. 25, Apr. 29, 2020.
[55] Kacha, *supra* note 54.
[56] *Algerian state tightens screws on online media*, France 24 (Dec. 18, 2020), https://www.france24.com/en/live-news/20201218-algerian-state-tightens-screws-on-online-media.
[57] Khaled Moulay Idriss, *Ma'reket qawanīn al-nashr 'ala al-internet fī mūrītanīa*, al-'Arabī al-Safīr (Jun. 8, 2020), https://assafirarabi.com/ar/32154/2020/07/08/%D9%85%D8%B9%D8%B1%D9%83%D8%A9-%D9%82%D9%88%D8%A7%D9%86%D9%8A%D9%86-%D8%A7%D9%84%D9%86%D8%B4%D8%B1-

making one believe destruction, disaster, or sabotage… or a supplies shortage will occur."[58] Because the Mauritanian government can judge what they believe to be false, the law could target legitimate statements of concern regarding crises or shortages. The law punishes violators with imprisonment ranging from three months to five years and fines amounting to between $150 and $600, which commentators labeled an especially harsh punishment.[59] The Law on Combating Manipulation of Information ostensibly seeks to prevent false information from spreading "especially during periods of elections and during health crises."[60] However, the law allows for enforcement during elections and under undefined "exceptional circumstances," inviting concern about its scope.[61]

On March 17, 2020, King Abdullah of Jordan approved a state of emergency under Article 124 of the Constitution activating Defense Law No. 13 of 1992, which authorizes the prime minister to restrict basic rights. The declaration empowered the prime minister to "deal firmly with anyone who spreads rumors, fabrications, and false news that sows panic about COVID-19.[62] Then, pursuant to the emergency declaration, King Abdullah issued Defense Order No. 8 on April 15, 2020, prohibiting "publishing, re-publishing, or circulating any news about the epidemic in order to terrify people or cause panic among them."[63] The order specifies penalties of up to three years in prison and a fine equivalent to $4,230.[64]

%D8%B9%D9%84%D9%89-%D8%A7%D9%84%D8%A5%D9%86%D8%AA%D8%B1%D9%86%D8%AA-%D9%81%D9%8A-%D9%85%D9%88%D8%B1/.

[58] *Id*.

[59] *Id; Mauritania Passes Law On False Publication Amidst Crackdown On Online Dissent*, All Africa (Aug. 10, 2020), https://allafrica.com/stories/202008100832.html.

[60] *Mauritania: President Should Lead Reform Process*, Hum. Rts. Watch (Aug. 1, 2020), https://www.hrw.org/news/2020/08/01/mauritania-president-should-lead-reform-process.

[61] Idriss, *supra* note 57.

[62] Wiseman, *supra* note 45.

[63] Prime Ministry of Jordan (@PrimeMinistry), Twitter (Apr. 15, 2020), https://twitter.com/PrimeMinistry/status/1250456883036672000.

[64] *Jordan: Free Speech Threats Under Covid-19 Response*, Hum. Rts. Watch (May 5, 2020), https://www.hrw.org/news/2020/05/05/jordan-free-speech-threats-under-covid-19-response.

Other executives similarly enacted emergency measures that include disinformation provisions. In Palestine, Prime Minister Mahmoud Abbas enacted Emergency Order No. 1 of 2020 on March 6. The order "bann[ed] dealing with, broadcasting, or sharing any rumors or false information" and required people to only use information from official sources.[65] In Sudan, Prime Minister Abdallah Hamdok passed Emergency Order No. 1 of 2020, Article 5 of which criminalizes "Publishing incorrect data or information, including rumours via any of the means of publishing or misleading the authorities about the coronavirus."[66] Although these measures would ideally lapse with the end of the COVID-19 pandemic, MENA governments have often seized authority during states of exception only to maintain those powers long after a crisis subsides.[67] Each measure leaves authorities to determine whether statements are true and interpret broadly defined terms.

Elsewhere, authorities applied existing penalties to COVID-19 disinformation. On April 18, 2020, Emirati authorities announced fines of up to the equivalent of $5,445 for people who share medical information about coronavirus that contradicts official statements.[68] Omani authorities publicized that they would enforce the country's 2011 cybercrime law to target false news regarding COVID-19.[69] Saudi Arabia took a similar approach, declaring that those who spread COVID-19 rumors or misinformation online would face up to five years imprisonment and fines of up to $800,000 pursuant to the kingdom's cybercrime law.[70] Egyptian authorities

[65] Emergency Order No. 1 of 2020, *al-Waqā'i'al-Falastīnīyah*, no. 21, Mar. 25, 2020.
[66] Sudan PM Hamdok issues first Emergency Order for 2020, Dabanga (Apr. 13, 2020), https://www.dabangasudan.org/en/all-news/article/sudan-pm-hamdok-issues-first-emergency-order-for-2020.
[67] *See generally* Giorgio Agamben, State of Exception (2003).
[68] Wiseman, *supra* note 45.
[69] Kabeer Yousuf, *Spreading rumors, fake news punishable by law*, Oman Observer (Feb. 14, 2021), https://www.omanobserver.om/article/3501/Main/spreading-rumours-fake-news-punishable-by-law.
[70] Khitam Al Amir, *Coronavirus: Saudi Arabia says rumour mongers will be jailed, fined*, Gulf News (Mar. 3, 2020), https://gulfnews.com/world/gulf/saudi/coronavirus-saudi-arabia-says-rumour-mongers-will-be-jailed-fined-1.70124319.

announced that those who spread rumors about COVID-19 would face five years in jail and fines up to $1,270.[71]

Restrictions on speech amid a pandemic do not necessarily violate international human rights law, which allows for restrictions on freedom of expression that meet the requirements of necessity and proportionality. However, the overbroad construction of MENA disinformation statutes raises concerns about whether restrictions fulfill international standards. Examining the enforcement of such measures redoubles those concerns, showing that MENA disinformation measures are used as tools of repression.

C. Enforcement of Disinformation Legislation for Repression

Even before the COVID-19 pandemic, some governments established a record of arresting activists and journalists pursuant to disinformation laws, including those passed since the Arab uprisings and pre-existing measures newly instrumentalized for restricting speech.

Two examples from the Persian Gulf region demonstrate this trend. Bahraini human rights activist Nabeel Rajab was arrested in June 2016 for "disseminating false rumors in time of war" for tweets criticizing Bahrain's role in the war in Yemen.[72] After Rajab wrote a New York Times op-ed from prison about human rights abuses in Bahrain, he faced additional charges of "false news and statements and malicious rumors that undermine the prestige of the kingdom."[73] In May 2018, an Emirati court convicted Ahmed Mansoor of using his social media accounts to publish

[71] *Jail term and EGP 20,000 fine for spreading rumours about coronavirus: Egypt's prosecution*, Ahram Online (Mar. 28, 2020), https://english.ahram.org.eg/NewsContent/1/64/366161/Egypt/Politics-/Jail-term-and-EGP-,-fine-for-spreading-rumours-abo.aspx.
[72] *Bahrain: Activist Nabeel Rajab jailed for 'fake news'*, BBC News (Jul. 10, 2017), https://www.bbc.com/news/world-middle-east-40558063.
[73] *Bahrain activist charged after letter in New York Times*, Associated Press (Sept. 5, 2016), https://apnews.com/article/6368ffafbb29453aaf154bf9666acf81.

false information, sentencing him to ten years in prison.[74] An early adopter of anti-disinformation measures soon after the Arab uprisings, the U.A.E. charged Mansoor under a 2012 cybercrimes law.[75] Called the U.A.E.'s "most celebrated human rights activist," Mansoor faced six charges related to his activism.[76]

The case of Egypt exemplifies how a new wave of fake news laws passed since 2015 has been instrumentalized for repression. The Egyptian government, led by President Abdel Fattah el-Sisi, has used its 2018 cybercrime law as a cudgel against activists and journalists. In August 2020, a terrorism court sentenced Bahey el-Din Hassan, a prominent human rights defender, to fifteen years in prison for "disseminating false news" under the cybercrime law.[77] In June 2021, Egypt's Public Prosecutor charged the prominent human rights activist Hossam Bahgat in case number 35/2020 with, *inter alia*, publishing false news under Law 175/2018 and Article 188 of the Penal Code.[78] The Public Prosecutor based the charge on a December 2020 tweet in which Bahgat accused a former president of the National Election Authority of fraud.[79]

Since the COVID-19 pandemic began, many Arab States have escalated enforcement of anti-disinformation laws to ostensibly discourage spreading unreliable information about the disease. In the early weeks of the COVID-19 pandemic, Moroccan authorities arrested at least a

---

[74] *UAE rights activist Ahmed Mansoor sentenced to 10 years in prison*, Al Jazeera (May 30, 2018), https://www.aljazeera.com/news/2018/5/30/uae-rights-activist-ahmed-mansoor-sentenced-to-10-years-in-prison.
[75] *Id*.
[76] *The Persecution of Ahmed Mansoor*, Hum. Rts. Watch (Jan. 27, 2021), https://www.hrw.org/report/2021/01/27/persecution-ahmed-mansoor/how-united-arab-emirates-silenced-its-most-famous-human.
[77] *Egypt: Human rights defender Bahey el-Din Hassan handed outrageous 15-year prison sentence*, Amnesty Int'l (Aug. 25, 2020), https://www.amnesty.org/en/latest/press-release/2020/08/egypt-human-rights-defender-bahey-eldin-hassan-handed-outrageous-15-year-prison-sentence/.
[78] *Egyptian journalist Hossam Bahgat is set to go on trial for a tweet*, Comm. to Protect Journalists (Sep. 13, 2021), https://cpj.org/2021/09/egyptian-journalist-hossam-bahgat-is-set-to-go-on-trial-for-a-tweet/; *The Public Prosecution orders Hossam Bahgat's release and charges him with insulting the election commission, and publishing false news of electoral fraud*, Egyptian Initiative for Personal Rts. (June 16, 2021), https://eipr.org/en/press/2021/06/public-prosecution-orders-hossam-bahgat%E2%80%99s-release-and-charges-him-insulting-election.
[79] *Egyptian journalist Hossam Bahgat is set to go on trial for a tweet*, Comm. to Protect Journalists (Sept. 13, 2021), https://cpj.org/2021/09/egyptian-journalist-hossam-bahgat-is-set-to-go-on-trial-for-a-tweet/.

dozen people for spreading disinformation related to the public health emergency.[80] Jordanian police, too, have escalated arrests of individuals who allegedly spread false information about the pandemic around the country.[81] The Palestinian Authority arrested citizens who claimed online that COVID-19 had spread to their towns and neighborhoods.[82] In spring 2020, Saudi police conducted arrests across the kingdom pursuant to the provision of the country's cybercrime law regarding spreading rumors, invoking COVID-19 as a justification.[83] Although these cases do not draw the same scrutiny as measures taken against journalists or activists, they underscore key issues in MENA disinformation laws. In each case, governments claimed statements about COVID-19 were false because they did not comport with official narratives, but arrestees could have believed in the veracity of their statements. Leaving States as the deciders of truth allows for significant discretion in enforcement.

In addition to ordinary citizens, enforcement of disinformation and fake news laws during the pandemic targets journalists. Human rights law calls for enhanced scrutiny of the enforcement of restrictions on freedom of expression against journalism.[84] Invoking its COVID-19 measures, the Jordanian government arrested several journalists, media executives, and politicians accused of spreading disinformation about COVID-19 and government pandemic restrictions.[85] When Reuters reported that Iraqi authorities were hiding the severity of the country's COVID-19

[80] *Morocco makes a dozen arrests over coronavirus fake news*, Reuters (Mar. 19, 2020), https://www.reuters.com/article/us-health-coronavirus-morocco-idUSKBN2162DI.
[81] *See Dhabt marūj ishā'āt bi-sha'n nīyet al-hakūmah i'lān khatr shāmel*, AlRai (Mar. 26, 2020), http://alrai.com/article/10532916/.
[82] *PA arrests 4 for spreading 'fake news' on coronavirus*, Middle E. Monitor (Mar. 11, 2020), https://www.middleeastmonitor.com/20200311-pa-arrests-4-for-spreading-fake-news-on-coronavirus/.
[83] Iman Al-Khattaf, *Saudi Police Fight Rumor Promoters*, Asharq Al-Awsat (May 6, 2020), https://english.aawsat.com/home/article/2270396/saudi-police-fight-rumor-promoters.
[84] UNHRC, *supra* note 10.
[85] *Jordan: Free Speech Threats Under Covid-19 Response*, Hum. Rts. Watch (May 5, 2020), https://www.hrw.org/news/2020/05/05/jordan-free-speech-threats-under-covid-19-response.

outbreak, the Iraqi government labeled the news "deliberate misinformation,"[86] suspended Reuters' license, and fined the outlet.[87] Under the authority granted in the 2018 cybercrime law, on April 9, 2020, Egypt's Supreme Council for Media Regulation (SCMR) blocked the website of a new opposition publication that reported on conditions for political prisoners amid the COVID-19 pandemic.[88] The SCMR also blocked a dozen unspecified media websites, shut down social media pages, and expelled a Guardian reporter, all for allegedly spreading fake news about COVID-19.[89] In the first six months of the COVID-19 pandemic, Egyptian authorities arrested at least ten doctors and six journalists for criticism of government COVID-19 measures.[90] Although the long-term impact of these measures remains uncertain, early signs indicate that the enforcement of disinformation laws from COVID-19 will likely continue.

Vaguely constructed disinformation laws have provided MENA States with a justification to crack down on expression both before and during the COVID-19 pandemic. Based on regional precedents, enforcement of disinformation, fake news, and rumor-mongering provisions may become another tool of digital authoritarianism in the Middle East and North Africa. In many cases, they may transform from temporary emergency measures into permanent fixtures in States' arsenals of repression.

---

[86] Government of Iraq (@IraqiGovt), Twitter (Apr. 20, 2020), https://twitter.com/IraqiGovt/status/1245812546931695616.
[87] Alex MacDonald, *Coronavirus: Iraq suspends Reuters' licence after report on infection numbers*, Middle E. Eye (Apr. 3, 2020), https://www.middleeasteye.net/news/coronavirus-iraq-reuters-licence-suspended-after-report.
[88] *Egypt blocks Darb news website*, Comm. to Protect Journalists (Apr. 14, 2020), https://cpj.org/2020/04/egypt-blocks-darb-news-website/.
[89] *Egypt blocks online "fake news" about coronavirus*, Reporters without Borders (Apr. 3, 2020), https://rsf.org/en/news/egypt-blocks-online-fake-news-about-coronavirus.
[90] *Egypt arrests doctors, silences critics over virus outbreak*, Associated Press (Jul. 6, 2020), https://apnews.com/article/health-united-nations-ap-top-news-virus-outbreak-international-news-cf9528ebff1d5dd7e3b95d467d7e9418.

D. Successful Civil Society Opposition to Disinformation Legislation

Some States have withdrawn proposed disinformation laws after criticism from civil society organizations. These instances show the effectiveness of opposition when representative bodies introduce excessively restrictive disinformation laws. Despite the gains that digital authoritarianism has made, its contest with digital activism remains fluid.[91]

Morocco's Government Council approved Draft Law No. 22.20 on March 19, 2020. Article 16 "criminalized us[ing] social networks, open broadcast networks or similar networks to publish or promote electronic content containing false information."[92] The Draft Law imposed punishments of "imprisonment for three months to two years and a fine of 1,000 to 5,000 dirhams." Civil society organizations and opposition parties spoke out against the bill, expressing "concerns about the abuses of freedom in the management of the Covid-19 pandemic and... the increase in arrests for crimes of opinion."[93] Lawmakers suspended the Draft Law in May 2020.[94]

In Tunisia, Member of Parliament Mabrouk Karsheed introduced an anti-disinformation measure in March 2020 amid rising concerns about COVID-19.[95] Hours after news of the proposal

---

[91] Ahmed Shaheed & Benjamin Greenacre, *Binary Threat: How Governments' Cyber Laws and Practice Undermine Human Rights in the MENA Region*, Project on Middle E. Pol. Sci. 8 (August 2021), https://pomeps.org/wp-content/uploads/2021/08/POMEPS_Studies_43_Draft3-1.pdf

[92] *Morocco: Government must fully withdraw draft law on social media*, Article 19 (Jul. 10, 2020), https://www.article19.org/resources/morocco-social-media/.

[93] *Maroc: Non au choc liberticide en période de crise sanitaire*, Article 19 (Apr. 30, 2020), https://www-article19-org.translate.goog/fr/resources/maroc-non-au-choc-liberticide-en-periode-de-crise-sanitaire/?_x_tr_sl=auto&_x_tr_tl=auto&_x_tr_hl=en-US&_x_tr_pto=nui,op; *Draft law would violate right to freedom of expression in Morocco*, Euromed Human Rights Monitor (2020), https://euromedmonitor.org/uploads/reports/moroccoen.pdf

[94] *Freedom on the Net 2020: Morocco*, Freedom House (2020), https://freedomhouse.org/country/morocco/freedom-net/2020.

[95] Dina Samaro & Emna Sayadi, *Tunisia's Parliament on COVID-19: an initiative to fight disinformation or an opportunity to violate fundamental rights?*, Access Now (Apr. 1, 2020), https://www.accessnow.org/tunisias-parliament-on-covid-19-an-initiative-to-fight-disinformation-or-an-opportunity-to-violate-fundamental-rights/.

leaked, Karsheed withdrew the law amid opposition from civil society groups and the National Lawyers Association.[96]

Amid a national crackdown on dissent,[97] Iraq's Parliament reintroduced a 2011 cybercrimes bill in 2019.[98] Parliament considered the bill as Iraq faced some of its first COVID-19 cases in the spring of 2020. The bill included a prohibition on "publishing or broadcasting false or misleading events for the purpose of weakening confidence in the electronic financial system, electronic commercial or financial documents, or similar things, or damaging the national economy and financial confidence in the state."[99] Human rights organizations mobilized against the bill, and Parliament suspended it in February 2021.[100]

These examples demonstrate the important role that MENA civil society organizations play as watchdogs, especially during crises. However, speaking out early is crucial; the same disinformation measures that civil society organizations protest may limit those organizations' ability to criticize the government. Consulting with independent civil society organizations when drafting disinformation laws may help legislators narrowly construct provisions to limit concerns about unnecessarily restricting expression.

IV. Conclusion

In much of the Middle East and North Africa, governments have passed legislation to curb disinformation and fake news. After the Arab uprisings, a first wave of States passed cybercrime laws restricting online speech. Then, as fake news narratives gained traction during the Trump

[96] Omar Sattar, *Activists fear Iraqi cybercrime law could limit press freedoms*, Al Monitor (Dec. 2, 2020), https://www.al-monitor.com/originals/2020/12/iraq-parliament-cybercrime-freedom.html#ixzz79xqAd600.
[97] Alissa J. Rubin, *Iraq in Worst Political Crisis in Years as Death Toll Mounts From Protests*, N.Y. Times (Dec. 21, 2019), https://www.nytimes.com/2019/12/21/world/middleeast/Iraq-protests-Iran.html.
[98] *Iraq: Scrap Bill to Restrict Free Speech*, Hum. Rts. Watch (Nov. 25, 2020), https://www.hrw.org/news/2020/11/25/iraq-scrap-bill-restrict-free-speech.
[99] *Id.*
[100] *Id*.

administration, MENA governments passed a round of fake news laws. Amid the COVID-19 pandemic, States have enacted new measures criminalizing disinformation, rumor-mongering, and fake news. Although some States may eventually scale back enforcement, authoritarian MENA governments have typically used states of exception such as the COVID-19 pandemic as a one-way ratchet for State authority.

MENA disinformation laws often contain undefined or overbroad terms that invite significant government discretion in interpretation and enforcement. They often criminalize legitimate political speech and sharing information that is, or is believed to be, true. The scope of emergency measures is often indefinite, allowing for their extension far beyond the COVID-19 pandemic. Punishments typically include lengthy prison sentences and hefty fines. In many cases, MENA laws regarding disinformation do not comport with international human rights standards for freedom of expression because they do not meet the necessity and proportionality principles required for permissible restrictions on the freedom of expression.

The enforcement of those laws, which often targets political critics and journalists, reinforces concerns about the purpose of disinformation measures. However, a robust civil society can prevent the enactment of overly broad measures that restrict speech. As other states consider how to best address disinformation, the MENA should serve as a cautionary tale for how authoritarian governments can use disinformation laws for repression.

The State of Disinformation in Sub-Saharan Africa: A Case Study of Ethiopia

Introduction

Concerns over the spread of disinformation have dominated the media in recent years gaining traction globally in the aftermath of the Facebook-Cambridge Analytica scandal,[1] and more recently in the wake of the coronavirus pandemic with the spread of Covid-19 conspiracy theories on social media platforms.[2] In Africa, interest in disinformation became more pronounced in 2016 following news that Bell Pottinger, a British PR firm, ran disinformation campaigns to "stir up racial tensions" in South Africa in 2016 on behalf of the Gupta family as "a counter-narrative to growing backlash over the family's central role in grand corruption and state capture."[3] While fake news or disinformation in Africa is not new and dates as far back as "the colonial era in Zimbabwe,"[4] the increasing penetration of the Internet and social media in Sub-Saharan Africa

[1] *See* Larry Madowo, *Is Facebook Undermining Democracy in Africa?*, BBC NEWS (May 24, 2019), https://www.bbc.com/news/world-africa-48349671. "In 2018, Facebook and British data analytics firm Cambridge Analytica were at the centre of a dispute over the harvesting and use of personal data of more than 230 million users" with intent to alter voting behavior in multiple countries. While the United States was at the forefront of the discussion after Facebook confirmed it improperly accessed the personal data of 50 million of the network's users in connection to the 2016 Presidential elections that led Donald Trump to the presidency, Cambridge Analytica's activities also touched the African continent. Cambridge Analytica or its parent company SCL Group worked on the 2013 and 2017 campaigns of Kenya's President. The company was also hired to support the re-election bid of then-president Goodluck Jonathan of Nigeria in 2015.

[2] For a discussion on disinformation and conspiracy theories surrounding the coronavirus pandemic in Africa, see Diomma Dramé, *The Health Crisis: Fertile Ground for Disinformation*, UNESCO (Mar. 2020), https://en.unesco.org/courier/2020-3/health-crisis-fertile-ground-disinformation.

[3] David Segal, *How Bell Pottinger, P.R. Firm for Despots and* Rogues, Met Its End in South Africa, N.Y. TIMES (Feb. 4. 2018), https://www.nytimes.com/2018/02/04/business/bell-pottinger-guptas-zuma-south-africa.html. The Guptas are a wealthy family in South Africa that was deeply "intertwined" with former President Jacob Zuma's presidency that critics referred to them as the "Zupta regime." The family hired Bell Pottinger to "draw attention away from them" by engaging in "grass-roots political activism intended to help poor blacks." The firm was "accused of setting off racial tensions through a furtive campaign built on Twitter bots, hate-filled websites and speeches. All were pushing a highly toxic narrative, namely that whites in South Africa had seized resources and wealth while they deprived blacks of education and jobs."

[4] Admire Mare et. al., *"Fake News" and Cyber-Propaganda in Sub-Saharan Africa: Recentering the Research Agenda*, 40 AFRICAN JOURNALISM STUDIES 1, 2 (2019).

has created another avenue to further enable its dissemination.[5] This paper explores the landscape of disinformation in Sub-Saharan Africa and attempts by states to address it.

Part I explores how increasing access to the Internet and social media facilitates the spread of disinformation, discusses the primary actors in disinformation campaigns, and why disinformation is a problem in Sub-Saharan Africa. Part II addresses attempts by States such as Ethiopia to address disinformation using domestic legislation. This chapter argues that while disinformation poses detrimental challenges to many Sub-Saharan African countries including exacerbating ethnic conflicts, endangering peace and security, and spurring violence offline, creating legislation that effectively curbs disinformation while safeguarding human rights is harder in practice.

I. The Disinformation Landscape in Sub-Saharan Africa

A. Internet and Social Media Usage Facilitates the Disinformation Ecosystem in Sub-Saharan Africa.

Internet penetration has been historically low in Sub-Saharan Africa due to infrastructure gaps.[6] However, increasing mobile phone subscriptions across the continent has enabled wider access to the Internet and with that, greater access to social media platforms.[7] As of 2019, roughly 240 million people are mobile phone subscribers in Sub-Saharan Africa representing a penetration rate of forty-five percent,[8] while internet penetration stands at 29.7 percent.[9] Sub-Saharan Africa

[5] Richard Ngamita, *Charting the Link Between disinformation, Disruptions, Diseases and the Diaspora in Cameroon and DR Congo*, CIPESA (May 21, 2021), https://cipesa.org/2021/05/charting-the-link-between-disinformation-disruptions-diseases-and-the-diaspora-in-cameroon-and-dr-congo/.
[6] Angelo Attanasio et. al, *Internet: Africa Starts to Open its Window to the World*, ALJAZEERA (2020), https://interactive.aljazeera.com/aje/2016/connecting-africa-mobile-internet-solar/internet-connecting-africa.html
[7] *Id.*
[8] GSMA, *Mobile Internet Connectivity 2020 Sub-Saharan Africa Factsheet* (2020), https://www.gsma.com/r/wp-content/uploads/2020/09/Mobile-Internet-Connectivity-SSA-Fact-Sheet.pdf.
[9] WORLD BANK, INDIVIDUALS USING THE INTERNET – SUB-SAHARAN AFRICA (2019), https://data.worldbank.org/indicator/IT.NET.USER.ZS?locations=ZG.

is projected to remain the fastest growing region with mobile internet users expected to reach 483 million people by 2023.[10] As internet penetration continues to grow across Africa, so has the use of social media.[11] Roughly 76.54 percent Africans use Facebook as compared to 8.78 percent for Twitter.[12] Africans generally rely on a wide range of print and broadcast media sources of information, alongside online sources, to keep abreast of social, economic and political issues.[13] However, social media plays an increasingly prominent role for sharing news and local information partly because mainstream media has a history of being controlled and censored by governments in many countries.[14] For example, a study across fourteen African countries found that fifty-four percent of "young people read news on social media, and a third spend . . . [over] four hours a day online, mainly on their smartphones."[15] But, as more Africans gain access to the Internet and social media platforms, another avenue for disinformation campaigns has emerged.[16]

While fake news or disinformation in Africa is not new and dates as far back as "the colonial era in Zimbabwe, where it was deployed by the colonial state as a response to political paranoia and insecurity,"[17] the increasing penetration of the Internet and social media in Sub-Saharan Africa has further enabled its dissemination.[18] Recent studies show digital disinformation – that is disinformation spread over the internet including platforms such as Facebook, Twitter and

[10] GSMA, *The Mobile Economy Sub-Saharan Africa 2019*, GSM ASS'N (2019), https://data.gsmaintelligence.com/api-web/v2/research-file-download?id=45121567&file=2794-160719-ME-SSA.pdf.
[11] Bridget Boakye, *Social Media Futures: Changing the African Narrative*, TONY BLAIR INST. FOR GLOBAL CHANGE (Apr. 19, 2021), https://institute.global/policy/social-media-futures-changing-african-narrative.
[12] GlobalStates, *Social Media Stats Africa* (2020), https://gs.statcounter.com/social-media-stats/all/africa.
[13] Yinka Adegoke, *Why Social Media is the Only Media in Africa*, WORLD ECON. F. (Mar 6, 2017), https://www.weforum.org/agenda/2017/03/why-social-media-is-the-only-media-in-africa/.
[14] *Id.*
[15] Pauline Bax et. al., *Online Disinformation Campaigns Undermine African Elections*, BLOOMBERG NEWS (Oct. 13, 2020), https://www.bloomberg.com/news/articles/2020-10-13/disinformation-campaigns-on-facebook-twitter-google-undermine-african-election.
[16] Kwami Ahiabenu et. al., *Media Perspectives on Fake News in Ghana*, PEN PLUS BYTES (May 2, 2018), https://www.penplusbytes.org/wp-content/uploads/2018/05/report-HIGH-1.pdf.
[17] *Id.*
[18] Ngamita, *supra* note 5.

WhatsApp – is increasingly becoming a common feature of Africa's domestic landscape.[19] Whereas "[t]raditional media houses serve gatekeeping and watchdog roles,[20] curating information and content that is consumed by the public[,]. . . with the increase in the usage of social media, the floodgates have been opened for user-generated content that is not constrained by any editorial limitation."[21] This "means that all kinds of news and content, including fake news, are generated and distributed at the speed of light."[22]

Disinformation over the Internet and social media platforms takes two primary forms in Sub-Saharan Africa. First, credible websites or verified social media accounts post fake news with the awareness the information is false and/or misleading.[23] Second, fake news websites – which "clone other popular websites" to appear "credible" post false information intending to mislead or misinform people.[24] While the second form of disinformation garners more attention particularly during election periods in many African countries where bots and fake accounts are created to smear opposition candidates, both forms of disinformation are increasingly intertwined. For example, "prior to Uganda's January 2021 election, a network of inauthentic social media accounts operating Facebook, Instagram, and Twitter spread coordinated disinformation supporting the ruling party. Some of these accounts were directly operated by the Ugandan government through the Government Citizen Interaction Center . . . at the Ministry of Information and Communications Technology and National Guidance."[25] Several of the accounts were also linked to "the

[19] *See, e.g.*, Africa Center for Strategic Studies, *Domestic Disinformation on the Rise in Africa* (Oct. 6, 2021), https://africacenter.org/spotlight/domestic-disinformation-on-the-rise-in-africa/.
[20] It is important to note that, "'Fake news' has also found its way into the mainstream media [in the African content] largely through the weakening of gatekeeping infrastructure and overreliance on online sourcing practices and cultures." The authors argue that this is as a result of the "juniorization of newsrooms and recruitment of interns and correspondents have been identified as some of the reasons for the decline in quality news production." Mare, *supra* note 4, at 2.
[21] Ahiabenu, *supra* note 16, at 2.
[22] *Id.*
[23] *Id.* at 3.
[24] *Id.*
[25] Africa Center for Strategic Studies, *supra* note 19.

spokesperson" for the President's son, a Ugandan military "lieutenant general."[26] These accounts "spread disparaging disinformation about opponents through misleading images, claims, and hashtags—such as exploiting homophobic sentiment in Uganda by asserting that Bobi Wine [the opposing party front-liner] is gay."[27] The coordinated "inauthentic" messaging was "amplified" by "copying and pasting posts and hashtags across dozens of pages and groups created by the fake user accounts."[28] A Facebook page was created for a fake news outlet -- "Kampala Times," "which helped give the false content a veneer of legitimacy."[29]

B. Disinformation Poses Serious Challenges for African Countries.

With the proliferation of digital technologies across the continent, "challenges posed by viral disinformation have multiplied, with fake news stories from the bizarre to the dangerous becoming a ubiquitous feature of Africa's political landscape over recent years."[30] For example, in 2018, President Muhammadu Buhari, who has suffered from various health problems, found himself compelled to deny rumors that he had died and been replaced in ceremonies by a Sudanese clone.[31] At a more extreme level, disinformation can fuel existing ethnic conflicts. For example, the "killing of Haacaaluu Hundeessaa, a popular singer and Oromo rights activist in June 2020," led to "a spiral of ethnic conflict supercharged by online hate speech and incitements to violence which ultimately translated into real-world atrocities."[32] The conflict "culminat[ed] [into] . . . an

[26] *Id.*
[27] *Id.*
[28] *Id.*
[29] *Id.*
[30] Will Marshall, *Fake News, Soft Authoritarianism and Challenges to Digital Democracy in Africa*, GLOB. RISK INSIGHTS (Mar. 9, 2021), https://globalriskinsights.com/2021/03/fake-news-soft-authoritarianism-and-challenges-to-digital-democracy-in-africa/.
[31] Yomi Kazeem, *Nigeria's President Denies He's Dead and has been Replaced by An Imposter*, QUARTZ AFRICA (Dec. 3, 2018), https://qz.com/africa/1482148/nigeria-buhari-denies-hes-dead-or-jubril-sudan-clone/.
[32] Marshall, *supra* note 30.

outright military offensive by the Ethiopian armed forces against ethnic separatists in the northern Tigray region."[33]

Disinformation poses serious challenges for African countries – many of whom are already fragile democracies. False and misleading information about health crises including the COVID-19 pandemic, for example, can be life-threatening.[34] For example, when the Ebola virus first hit Nigeria in 2014, fake news claiming it could be treated by drinking and bathing in salt water went viral leading to hospitalizations and the deaths of at least two persons due to excessive consumption of salt water.[35]

Disinformation can also endanger democracy on the continent.[36] Disinformation campaigns appear to be rampant during election cycles in many African countries. Research shows that "while laws are in place in most African countries to restrict political advertising on traditional media, there isn't enough accountability for platforms such as Facebook."[37] The lack of accountability allows disinformation to dominate social media during elections, which can further destabilize countries already dealing with uncertainty around elections. For example, during DR Congo's 2018 elections, "[s]ponsored content from Google and Facebook falsely alleged that former President Joseph Kabila's surrogate, Emmanuel Ramazani Shadary, had won the elections. The ads were published before the official results announcement by the Electoral Commission,

---

[33] *Id.*
[34] Mare, *supra* note 4, at 2.
[35] Tonye Bakare. *Fake News in Nigeria: A Complex Problem*, GOETHE INSTITUT (Dec. 2020), https://www.goethe.de/ins/ng/en/kul/mag/22061927.html; *See also*, Wadi Ben-Hirki. *Disinformation and Social Media Regulation – The Nigerian Experience*, Strong Cities Network (Jan. 2021), https://strongcitiesnetwork.org/en/guest-article-disinformation-and-social-media-regulation-the-nigerian-experience/.
[36] World Health Org., *Fighting Misinformation in the Time of COVID-19, One Click at a Time*, WTO (Apr. 27, 2021), https://www.who.int/news-room/feature-stories/detail/fighting-misinformation-in-the-time-of-covid-19-one-click-at-a-time.
[37] Pauline Bax et. al., *Online Disinformation Campaigns Undermine African Elections*, BLOOMBERG NEWS (Oct. 13, 2020), https://www.bloomberg.com/news/articles/2020-10-13/disinformation-campaigns-on-facebook-twitter-google-undermine-african-election..

which had been delayed. There were internet shutdowns in key cities, which made it even harder for fact checkers to verify any information related to the elections."[38]

Disinformation can also exacerbate existing ethno-religious conflicts endangering peace and security. For example, in 2018, an image was circulated online of "a baby with open machete wounds across his head and jaw" alleging the act resulted from "Fulani Muslims . . . killing Christians" in Nigeria.[39] This image, which did not originate in Nigeria, led to offline violence with about 238 people killed.[40]

C. Who Are the Orchestrators of Disinformation Campaigns?

The disinformation landscape in Sub-Saharan Africa is a hodgepodge of government, foreign state, non-state actors, and civilian-led fake news campaigns. Many governments are active participants in disinformation campaigns in Africa and view social media as another avenue to influence public opinion. In Zimbabwe for example, "at the behest of the president, many new social media accounts, and what we call 'paid Twitter,' were created ahead of the 2018 election."[41] Their main purpose was to overshadow online protests, disrupt conversations and to stalk and harass popular online influencers and opinion leaders.[42]

Non-state actors including terrorist groups, anti-government groups, and other local interest groups are also active on social media platforms. For example, in Nigeria, groups such as Boko Haram, Islamic State West Africa Province, and the Indigenous People of Biafra are all

[38] Ngamita, *supra* note 5.
[39] Yemisi Adegoke et. al., *Like. Share. Kill. Nigerian Police Say False Information on Facebook is Killing People*, BBC NEWS (2020), https://www.bbc.co.uk/news/resources/idt-sh/nigeria_fake_news.
[40] *Id.*
[41] *Id.*
[42] *Id.*

active on social media, using it to attack opponents and recruit followers, with the latter group even using automated bots to amplify their online voice and demands.[43]

Additionally, recent studies show foreign governments are increasingly funding disinformation campaigns across Africa.[44] In 2019, for example, "Facebook announced they had banned Israeli 'political consulting group' Archimedes who had reached millions of people with political manipulation in Nigeria and other African countries during elections, masking as local news organizations."[45] A recent study also found that a "disinformation network in Sudan" was connected to Russian oligarch Yevgeny Prigozhin and his Kremlin-aligned Internet Research Agency.[46] "However, only 2 of the 30 pages in the network that Facebook removed were operated from Russia. Most were run from Sudan and may have been domestically operated by local Sudanese citizens contracted by Prigozhin."[47]

Finally, civilians are also curators and disseminators of fake news online. For example, in 2019, a video that circulated on Facebook and WhatsApp showed Hausa farmers sprinkling insecticide on their beans before they were transported to the southeast of the country.[48] The purpose of this was to preserve the produce from pests such as weevils during the long journey. However, interpretations and voice-overs of the video emerged that drew on longstanding tensions between members of the Hausa and Igbo ethnic groups, claiming that the farmers were sprinkling poison, not insecticide.

---

[43] *Id.*
[44] *See, e.g.*, Davey Alba et. al., R*ussia Tests New Disinformation Tests New Disinformation Tactics in Africa to Expand Influence*, N.Y. TIMES (Sep. 1, 2020), https://www.nytimes.com/2019/10/30/technology/russia-facebook-disinformation-africa.html.
[45] Flemming Splidsboel Hansen et. al., *Disinformation Goes South*, SCIENCE NORDIC (Nov. 05, 2019, 10: 55 AM), https://sciencenordic.com/election-racism-researchers-zone/disinformation-goes-south/1588373.
[46] Africa Center for Strategic Studies, *Domestic Disinformation on the Rise in Africa* (Oct. 6, 2021), https://africacenter.org/spotlight/domestic-disinformation-on-the-rise-in-africa/.
[47] *Id.*
[48] Idayat Hassan & Jamie Hitchen, *Nigeria's Disinformation Landscape*, SOC. SCI. RSCH. COUNCIL (Oct. 6, 2020), https://items.ssrc.org/disinformation-democracy-and-conflict-prevention/nigerias-disinformation-landscape/.

II. Addressing Disinformation in Africa

With on-going impacts of disinformation felt across the African continent, African States agree that there is a need to address it, but what constitutes an effective solution differs among states.[49] Some governments have taken the approach of restricting access to social media as has been the practice with other traditional forms of media in "anti-democratic" countries.[50] For example, since 2015, roughly thirty African countries have restricted social media access.[51] In most countries, these restrictions occur around elections and political protests.[52] This approach raises human rights concerns as it can be used to "conceal human rights violations" particularly during periods of protest.[53]

Other governments have relied on existing legislations or created new laws to regulate speech more broadly in an attempt to curb disinformation. There are at least twenty-six countries in Africa with regulations on speech.[54] These regulations have been broadly criticized as restricting freedom of speech and as means of suppressing dissent.[55]

Further, some governments have also proposed specific legislation or passed laws targeting disinformation.[56] However, while most of the laws are intended to address disinformation, the line between the intentional sharing of misleading information (disinformation) and the sharing of

---

[49] Jeffery Conroy-Krutz, *African Governments Are Cracking Down on the News Media. Their Citizens Might Be Okay With That*, AFRO BAROMETER (May 14, 2019), https://afrobarometer.org/blogs/african-governments-are-cracking-down-news-media-their-citizens-might-be-okay

[50] Carlos Mureithi, *These are the African Countries that have Restricted Social Media Access*, QUARTZ AFRICA (Aug. 9, 2021), https://qz.com/africa/2044586/african-countries-that-have-restricted-social-media-access/.

[51] *Id.* In 2021 alone, "at least four African countries" have restricted access to social media -- Nigeria, Uganda, Senegal, and the Democratic Republic of Congo. *Id.*

[52] Yinka Adegoke, *Why Social Media is the Only Media in Africa*, QUARTZ AFRICA (Mar. 06, 2017), https://www.weforum.org/agenda/2017/03/why-social-media-is-the-only-media-in-africa/.

[53] *Id.*

[54] Global Partners Digital Limited, *Disinformation Tracker* (2021), https://www.disinformationtracker.org.

[55] *Id.*

[56] Alana Schetzer, *Governments Are Making Fake News A Crime – But it Could Stifle Free Speech*, THE CONVERSATION (Jul. 7, 2019), https://theconversation.com/governments-are-making-fake-news-a-crime-but-it-could-stifle-free-speech-117654.

misleading information with a lack of awareness that the shared information is actually misleading (misinformation) is often blurred. For example, under Cameroon's *Law Relating to Cyber Security and Cyber Criminality*, it is an offense to publish and propagate information online "without being able to attest to its veracity" or truthfulness.[57] Additionally, the laws typically have overly broad definitions of disinformation that, as critics opine, leave room for "arbitrary" application and human rights abuses.[58] For example, Nigeria's proposed disinformation bill, the *Protection from Internet Falsehoods and Manipulation and Other Related Matters Bill 2019*, states that individuals who transmit statements that authorities determine to be "false," likely to "influence the outcome of an election," or "prejudicial to the security of Nigeria," may be imprisoned for up to three years or fined up to 300,000 naira (US$844) or both.[59] Critics of the bill – including Human Rights Watch – claim it is too wide-ranging, representing a threat to free speech and legitimate political debate.[60]

Finally, it should be noted that some governments' discussions over regulating disinformation are often veiled attempts to silence dissent.[61] For example, during the #EndSars protests over police brutality in Nigeria, many activists used social media to call thousands of Nigerians to action. This led to calls by state governors for strict supervision and censorship of social media to thwart "subversive actions" and "avoid the spread of fake news."[62] Nigeria's president, Muhammadu Buhari, also complained that his government's critics were spreading

[57] Ngamita, *supra* note 5.
[58] Tafi Mhaka, *How Social Media Regulations Are Silencing Dissent in Africa*, ALJAZEERA (Nov. 12, 2020), https://www.aljazeera.com/opinions/2020/11/12/how-social-media-regulations-are-silencing-dissent-in-africa.
[59] Danielle Paquette, *Nigeria's 'Fake News' bill Could Jail People for Lying on Social Media. Critics Call it Censorship*, WASH. POST (Nov. 25, 2019), https://www.washingtonpost.com/world/africa/nigerias-fake-news-bill-could-jail-people-for-lying-on-social-media-critics-call-it-censorship/2019/11/25/ccf33c54-0f81-11ea-a533-90a7becf7713_story.html.
[60] *Id.*
[61] Adrian Shahbaz et. al., *The Global Drive to Control Big Tech*, FREEDOM HOUSE (2021), https://freedomhouse.org/report/freedom-net/2021/global-drive-control-big-tech
[62] *Id.*

"deliberate falsehoods and misinformation" through social media, claiming "that his government is oblivious of the pains and plights of its citizens."[63]

While state concerns over disinformation are legitimate, striking the proper balance between addressing it while safeguarding fundamental freedoms of speech and other human rights has proven a challenge for many governments, Ethiopia being a prime example.

A. A Case Study of Ethiopia's Disinformation Law

In March 2020, Ethiopia enacted the *Hate Speech and Disinformation Prevention and Suppression Proclamation* to address hate speech and disinformation.[64] Disinformation is defined in the *Proclamation* as "speech that is false, is disseminated by a person who knew or should reasonably have known the falsity of the information and is highly likely to cause a public disturbance, riot, violence or conflict."[65] The law provides that "any person who commits acts proscribed under [the Act] . . . shall be punished with simple imprisonment not exceeding two years or a fine not exceeding 100,000 birr [USD 2,907]."[66] The law also requires companies to "limit content flow" that is deemed to contravene the *Proclamation* by removing it from their platforms within 24 hours.[67] The stated purpose of the *Proclamation* is to "prevent and suppress by law the deliberate dissemination of hate speech and disinformation" because it "pose[s]" a threat to "social harmony, political stability, national unity, human dignity, diversity and equality."[68] Yet

[63] *Id.*
[64] Federal Negarit Gazette of the Federal Democratic Republic of Ethiopia, https://www.article19.org/wp-content/uploads/2021/01/Hate-Speech-and-Disinformation-Prevention-and-Suppression-Proclamation.pdf.
[65] *Id.* at 12341.
[66] *Id.*
[67] *Id.*
[68] *Id.*

"despite this articulation of a legitimate aim," the Proclamation has been criticized as "granting [the] government the authority to restrict a wide range of speech."[69]

During the drafting of the *Proclamation*, the government and the legislative body engaged scholars, journalists, civil society groups and political party leaders for feedback on the draft law.[70] There were at least three consultations held that gathered different stakeholders, including Access Now and Article 19 to comment on the legislation both orally and in writing.[71] Through these consultations, changes were made to the draft legislation. For instance, the definitions of terms such as disinformation was included.[72] The government also clarified that "speech will not be considered as disinformation and prohibited if a reasonable effort has been made under the circumstances by the person making the speech to ensure the veracity of the speech or if the speech is more inclined to political commentary and critique instead of being a factual or news report."[73] Finally, the government also reduced prison terms and gave courts more room to adjust penalties according to the offense.[74]

A recent study which reviews various approaches to legislating disinformation outlines four standards for "effective disinformation regulation."[75] First, regulations must "target the

[69] *See, e.g.*, Article 19, *Ethiopia: Hate Speech and Disinformation Law Must not Be Used To Suppress the Criticism of the Government,* (Jan. 19, 2021), https://www.article19.org/resources/ethiopia-hate-speech-and-disinformation-law-must-not-be-used-to-supress-the-criticism-of-the-government/.
[70] Ayele Addis Ambelu, *Hate Speech and Disinformation Prevention and Suppression Law in Ethiopia*, AFRICA NEWS CHANNEL (Oct. 25, 2021), https://www.africanewschannel.org/news/hate-speech-and-disinformation-prevention-and-suppression-law-in-ethiopia/
[71] *See, e.g.*, Article 19, *Ethiopia: Hate Speech and Disinformation Law Must not Be Used To Suppress the Criticism of the Government*, (Jan. 19, 2021), https://www.article19.org/resources/ethiopia-hate-speech-and-disinformation-law-must-not-be-used-to-supress-the-criticism-of-the-government/. Access Now and Article 19 are prominent human rights organizations.
[72] *Id.*
[73] Federal Negarit, *Gazette of the Federal Democratic Republic of Ethiopia*, https://www.article19.org/wp-content/uploads/2021/01/Hate-Speech-and-Disinformation-Prevention-and-Suppression-Proclamation.pdf.at 12342.
[74] https://www.accessnow.org/cms/assets/uploads/2020/05/ParlAmendComment.pdf.
[75] Ben Epstein, *Why It Is So Difficult to Regulate Disinformation Online*, *in* The Disinformation Age 1, 191–192 (Cambridge ed., 2020), https://www.cambridge.org/core/books/disinformation-age/why-it-is-so-difficult-to-regulate-disinformation-online/A7613D7394F18AAE8F241894E8DA064A

negative effects of disinformation while consciously minimizing any additional harm caused by the regulation itself."[76] Second, "regulations[s] must be proportional to the harm caused" and "powerful enough to cause change."[77] Regulations must also "be nimble, and better able to adapt to changes in technology and disinformation strategies than previous communication regulations."[78] Finally, "regulations should be as independent as possible from political leaders and leadership of the dominant social media and internet companies and guided by ongoing research in th[e] field."[79] Viewed as a standalone or weighed against these standards, Ethiopia's *Proclamation* is "flawed" in ways that hamper human rights.[80]

First, the "requirement in Ethiopia's law for reasonable knowledge that the information disseminated is false directly has a chilling effect on freedom of expression online." [81] Varying educational and digital literacy levels affect what constitutes "reasonable knowledge" of falsity and as such the ability to distinguish false information from accurate information.[82] Additionally, the requirement limits the ability of various groups to engage with "certain information" on social media platforms that they may disagree with due to fears of criminal liability.[83] There is a difference between false and misleading information spread by Russian troll farms meant to influence . . . election[s], and satirical articles" and Ethiopia's law does not make that distinction as clear as possible. [84]

Further, the law "unfairly presumes that one who sends such information and not necessarily the author should be held answerable" and also requires social media companies to

[76] *Id.*
[77] *Id.*
[78] *Id.*
[79] *Id.*
[80] *Id.*
[81] *Id.*
[82] *Id.*
[83] *Id.*
[84] *Id.* at 192–193.

"limit content flow," "hold[s] intermediaries liable for content policing," which "raises due process concerns," and "contravenes international human rights instruments."[85] Disinformation "can take many forms and is linked to a varied group of actors who create it, and a variety of platforms which are used to disseminate it, [h]owever, disinformation [that has the potential to be harmful] is always perpetuated on purpose by a particular group of responsible actors."[86] Effective regulations should be flexible to account for the different forms and how they may be employed across different platforms. While Ethiopia's law applies too broad a scope to what constitutes disinformation, it essentially leaves no room to account for the different forms of disinformation, the different actors engaged in it, the platforms used and how they may even change overtime. Put simply, the law effectively criminalizes forms of disinformation which may not pose any *grave* harm. For example, in 2020 amid the covid-19 pandemic, a journalist was accused of terrorism under the law for posting on Facebook that the government was "readying graves that could accommodate 200,000 corpses."[87]

Finally, while Ethiopia's approach of engaging different stakeholders in the drafting process should be lauded, it also serves as a cautionary tale that while a multi-stakeholder approach is crucial to crafting legislation that balances competing needs and human rights norms to address disinformation, it does not guarantee the outcome. Effectively legislating disinformation will require more than a mapping of who is at the table during the drafting process and whether their views are reflected in the final product. It also requires an acknowledgement that views and needs change over time. The process of crafting disinformation laws should always build in opportunities

[85] *Id.*
[86] *Id.* at 194.
[87] Elias Meseret, *Hate Speech and Disinformation Concerns Escalate in Ethiopia*, DEVEX (May 6, 2020), https://www.devex.com/news/hate-speech-and-disinformation-concerns-escalate-in-ethiopia-97095.

to test various iterations of the legislation in practice for specified periods of time, their impact on the regulated, and also leave room for data-based changes as necessary.

Conclusion

Disinformation remains a major problem across countries in Sub-Saharan Africa. While States have attempted to address it through regulations, the regulations are often loosely-defined and produce a "chilling effect" on human rights. Countries like Ethiopia recognize the need to engage different stakeholders to ensure that laws regulating disinformation do not violate human rights. However, striking the proper balance in reality remains a pipe dream – at least for now.

Legal and Regulatory Solutions

Disinformation challenges have driven a wide variety of actors to seek out solutions around the world. Proposed solutions to the problems posed by disinformation approach the phenomenon from a range of perspectives and angles. While some actors have applied preexisting legal and regulatory frameworks to disinformation, others have drafted new laws and rules purpose-made for disinformation. Solving the variety of problems associated with disinformation has thus far proven elusive, and current approaches taken by private entities, governments, and international institutions each have significant shortcomings.

This Section examines potential legal and regulatory solutions to address disinformation challenges. The essays contained in this Section exemplify the variety of options available, ranging from social media companies creating institutions to govern expression in public fora to international legal institutions taking up disinformation cases. The essays exemplify the difficulties present in creating sufficient legal solutions to the many dimensions of disinformation issues and further demonstrate the technical, procedural, and political challenges that regulators face.

When Big Tech Steps Up: The Role of Private Actors in Mitigating Misinformation

I. Introduction

Today, laws governing the dissemination of information online are murky and hotly contested. Traditionally, American laws around online information governance have been fairly permissive per the First Amendment of the U.S. Constitution. However, as the need to address harmful online speech becomes increasingly urgent, a move away from such permissibility may be warranted. Nevertheless, the growing role of popular tech platforms as public fora for political discussion means that any restrictions for speech should not be taken lightly. Evidence of these competing priorities can be found in the ongoing debate[1] about Section 230 of the 1996 Communications Decency Act. Section 230 provides immunity to website platforms with respect to their users' content. Thus, if this immunity were cabined or repealed altogether, tech companies could become legally liable for misinformation posted online.

Governing information online is further complicated by the global adoption of popular technology platforms like Facebook and Twitter.[2] Developed in accordance with American legal principles, the platforms are now tasked with navigating regulatory regimes outside of the U.S. Therefore, when it comes to determining which types of content should be shared on these

[1] In May 2020, President Trump issued an executive order intended to limit Section 230 protections on grounds that the law enables tech companies to unfairly censor conversative speech. *See* Anshu Siripurapu, *Trump and Section 230: What to Know*, COUNCIL FOR FOREIGN RELATIONS (Dec. 2, 2020), https://www.cfr.org/in-brief/trump-and-section-230-what-know. More recently, Senator Amy Klobuchar introduced legislation that would subject tech companies to greater liability when their users post harmful or deceptive communications on their platforms. *See* Health Misinformation Act of 2021, S.R. 21778, 117th Cong. § 1 (2021).

[2] Together, the social media platforms Facebook, YouTube, and Twitter have billions of monthly users around the world. *See* Simon Kemp, *Digital 2021 April Global Statshot Report*, DATA PORTAL (Apr. 12, 2021), https://datareportal.com/reports/digital-2021-april-global-statshot.

platforms, achieving consensus is nearly impossible. This struggle is exemplified in attempts to enforce domestic judgments on content removal across borders. *Glawischnig-Piesczek v. Facebook Ireland Ltd.,*[3] for instance, raised extraterritoriality questions when Austrian Green Party politician Eva Glawischnig-Piesczek attempted to enforce an Austrian judgment to remove defamatory Facebook posts outside of Austria. When Facebook declined to remove the posts from non-Austrian feeds, Glawischnig-Piesczek appealed to the European Court of Justice ("ECJ"), which affirmed Facebook's duty to remove defamatory posts regardless of where they are posted or accessed. In contrast, the ECJ sided with Google in *Google v. CNIL*,[4] finding that the tech company had no obligation to apply the European Union's right to be forgotten laws universally. In short, the Court confined the search engine's duty to comply with European citizens' content removal requests to the EU. In some sense, the ruling was a win for tech companies asserting that no one jurisdiction should be able to impose content removal mandates universally.

As the law attempts to keep pace with information technology, tech companies can play a significant role in mitigating misinformation on their platforms. Major tech companies like Facebook and Twitter have attempted to curb misinformation on their platforms by implementing their own regulatory regimes. These regimes, much like their state-sanctioned counterparts, include codes of conduct, enforcement mechanisms, and penalties for violations. Facebook has even attempted to create its own version of a supreme court, the Facebook Oversight Board, which has jurisdiction to review appeals of Facebook decisions to take down posts, deplatform user accounts, and offer policy recommendations.[5] A key difference from public governance, however,

[3] *Eva Glawischnig-Piesczek v. Facebook Ireland Limited*, Case C-18/18 ECLI:EU:C:2019:821.
[4] *Google LLC v. CNIL*, C-507/17, 2019 EUR-Lex CELEX No. 62017CJ0507 (Sept. 24, 2019).
[5] While Facebook's Oversight Board is a fascinating example of a private regime for information governance, further analysis is beyond the scope of this section. *See* Facebook Oversight Board, FACEBOOK, https://oversightboard.com/.

is tech companies' creative use of technology to root out violations. For instance, special algorithms and direct user feedback, among other tools, can be used to identify misinformation in real time.

The result of tech companies' efforts have found mixed success, due in large part to fragmented legal standards for online speech. It is hard to say whether tech companies will ever be able to generate universally applicable content rules in a world of abundant cultural and legal norms around freedom of speech. This paper attempts to frame this challenge by (1) outlining Facebook's and Twitter's policies on misinformation, (2) assessing the efficacy of these tools within the context of two case studies, and (3) analyzing the legal implications of each case study.

II. Overview of Private Sector Information Governance

Recent years have seen a proliferation of internal rules and tools at tech companies attempting to both curb and detect misinformation, including removing content connected with real-world violence, labeling deceptive content, and soliciting input from external fact-checkers and other users.[6] Although private companies have long had the capacity to "block, filter, mute, and decrease the visibility of online expression,"[7] many have harnessed new techniques to curb the spread of misinformation. For instance, in response to the spread of misinformation related to COVID-19, some tech companies have added fact-check labels to politicians' posts, disabled accounts attempting to sow discord, and prioritized posts of various public health agencies.[8]

---

[6] *See e.g.*, Rebecca Hamilton, *Governing the Global Public Square*, 62 Harv. Int'l L.J. 117, 2021; *Twitter Adds Warnings to Trump and White House Tweets, Fueling Tensions,* N.Y. TIMES (June 3, 2020), https://www.nytimes.com/2020/05/29/technology/trump-twitter-minneapolis-george-floyd.html.

[7] Danielle K. Citron & Neil M. Richards, *Four Principles for Digital Expression*, 95 Wash. U. L. Rev. 1353, 1356 (2018).

[8] *See* Davey Alba, *Virus Conspiracists Elevate a New Champion*, N.Y. TIMES (May 9, 2020); https://www.nytimes.com/2020/05/09/technology/plandemic-judy-mikovitz-coronavirus-disinformation.html; Raymond Zhong, et al., *Behind China's Twitter Campaign, a Murky Supporting Chorus*, N.Y. TIMES (June 8, 2020), https://www.nytimes.com/2020/06/08/technology/china-twitter-disinformation.html.

Both Facebook and Twitter strive to address misinformation while preserving freedom of expression by restricting their own power to remove content and soliciting input from external parties. Facebook claims that its Community Standards are intended to "create a place for expression and give people a voice"[9] while Twitter seeks to "serve the public conversation."[10] Accordingly, the companies will refrain from removing objectionable content if the public interest in maintaining it outweighs the risks of harm. When they must limit expression, the companies purport to rely on international human rights standards to preserve authenticity, safety, privacy, and dignity.[11] Although Facebook is reluctant to remove news on the grounds of falsity only, it will look to other characteristics as justification for removal, such as spam, hate speech, and fake accounts.[12] Twitter, on the other hand, will consider whether content has been posted with the intent to deceive. Deceptive content may include substantial edits to media composition, sequence, timing, or framing; additions and removals to auditory or visual information; and fabrications and simulations of media concerning real people.[13]

In line with efforts to minimize content removal to the greatest extent possible, Facebook and Twitter solicit third-party feedback to add labels and warnings to potentially deceptive content. Facebook has partnered with fact-checkers from the International Fact-Checkers network, whose role is to "identify, review, and take action" on "viral misinformation, particularly clear hoaxes

[9] Facebook Community Standards, FACEBOOK TRANSPARENCY CENTER, https://transparency.fb.com/policies/community-standards/, (last visited Oct. 27, 2021).
[10] The Twitter Rules, TWITTER HELP CENTER, https://help.twitter.com/en/rules-and-policies/twitter-rules, (last visited Oct. 27, 2021).
[11] Miranda Sissons, *Our Commitment to Human Rights,* FACEBOOK NEWSROOM (Mar. 16, 2021), https://about.fb.com/news/2021/03/our-commitment-to-human-rights/; *Defending and respecting the rights of people using our service*, TWITTER HELP CENTER, https://help.twitter.com/en/rules-and-policies/defending-and-respecting-our-users-voice (last visited Oct. 27, 2021).
[12] *Hard Questions: What's Facebook's Strategy for Stopping False News?*, FACEBOOK NEWSROOM (May 23, 2018), https://about.fb.com/news/2018/05/hard-questions-false-news/.
[13] Synthetic and manipulated media policy, TWITTER HELP CENTER, https://help.twitter.com/en/rules-and-policies/manipulated-media, (last visited Oct. 27, 2021).

that have no basis in fact."[14] Fact-checkers reviews may result in the addition of warning labels or more severe penalties, depending on the offender's record. Twitter's labeling system also relies on some external input. Specifically, Twitter consults with its Trust & Safety Council to ensure consideration of "global perspectives around the changing nature of online speech, including how [its] rules are applied and interpreted in different cultural and social contexts.[15] Though Twitter, unlike Facebook, does not currently utilize external fact-checkers to review content, it has launched a pilot program called Birdwatch which allows users to identify potentially misleading information. This community-driven approach enables users to add notes to Tweets they believe are misleading, which will become globally visible if they achieve "consensus from a broad and diverse set of contributors."[16]

A. *Private Governance Case Studies and Assessment of Effectiveness*

Actors throughout the world have taken advantage of social media channels to spread misinformation. Throughout 2019 and 2020, the Chinese Communist Party ("CCP") leveraged Facebook and Twitter, among other platforms, to control the narrative of anti-government protests in Hong Kong. In Germany, right-wing extremists wielded social media channels to co-opt QAnon conspiracy theories for their own political purposes. Both Facebook and Twitter have stepped up to address the spread of misinformation in these cases by doubling down on old strategies while implementing new ones. Together, the incidents serve as useful case studies on private governance

[14] Fact-Checking on Facebook, FACEBOOK FOR BUSINESS HELP CENTER, https://www.facebook.com/business/help/2593586717571940?id=673052479947730, (last visited Oct. 27, 2021).
[15] Our approach to policy development and enforcement philosophy, TWITTER HELP CENTER, https://help.twitter.com/en/rules-and-policies/enforcement-philosophy (last visited Oct. 27, 2021).
[16] Keith Coleman, *Introducing Birdwatch, a community-based approach to misinformation*, TWITTER BLOG (Jan. 25, 2021), https://blog.twitter.com/en_us/topics/product/2021/introducing-birdwatch-a-community-based-approach-to-misinformation.

of misinformation because they demonstrate (1) the clash of tech company policies with domestic rules and (2) the interplay of public and private regulatory frameworks.

1. Case Study 1: Misinformation in the 2019-2020 Hong Kong Protests

The CCP is particularly adept at spreading misinformation due to its extensive jurisdiction over all national communication channels. As such, the CCP has been known to manipulate images and videos with the aim of stoking nationalist and anti-Western sentiments.[17] During the 2019-2020 Hong Kong protests following the government's introduction of new extradition legislation, the CCP leveraged Facebook and Twitter to negatively portray protestors.[18] These efforts included painting protestors as paid provocateurs and misrepresenting the sources of injuries sustained by aggressive law enforcement.[19] The CCP has also successfully removed social media posts that fail to align with its messaging on the new legislation.[20]

a) Misinformation Mitigation Strategies

In line with stated policies on information governance, Twitter and Facebook took steps to mitigate the Chinese state's deceptive portrayal of Hong Kong protestors. Facebook removed seven pages, 15,500 people, and three groups with a combined 2,200 members suspected of CCP-

---

[17] *See e.g.*, Erika Kinetz, *Army of fake fans boosts China's messaging on Twitter*, AP NEWS (May 28, 2021), https://apnews.com/article/asia-pacific-china-europe-middle-east-government-and-politics-62b13895aa6665ae4d887dcc8d196dfc; Marcel Schliebs, et. al, *China's Public Diplomacy Operations: Understanding Engagement and Inauthentic Amplification of PRC Diplomats on Facebook and Twitter*, PROGRAMME ON DEMOCRACY AND TECHNOLOGY (May 11, 2021), https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/127/2021/05/Chinas-Public-Diplomacy-Operations-Dem.Tech-Working-Paper-2021.1-4.pdf, https://www.bbc.com/news/world-asia-china-58062630.

[18] *See e.g.*, John Dotson, *Chinese Covert Social Media Propaganda and Disinformation Related to Hong Kong*, JAMESTOWN FOUNDATION (Sep. 6, 2019), https://jamestown.org/program/chinese-covert-social-media-propaganda-and-disinformation-related-to-hong-kong/; Grace Shao, *Social media has become a battleground in Hong Kong's protests*, CNBC (Aug. 16, 2019), https://www.cnbc.com/2019/08/16/social-media-has-become-a-battleground-in-hong-kongs-protests.html.

[19] *Id.*

[20] Since the CPP blocks both Facebook and Twitter in mainland China, the efforts were likely designed with the aim of influencing perceptions of the protests overseas. *See* Louise Matsaki, *China Attacks Hong Kong Protesters With Fake Social Posts*, WIRED (Aug. 19, 2019), https://www.wired.com/story/china-twitter-facebook-hong-kong-protests-disinformation/.

coordinated inauthentic behavior.[21] Twitter also took down accounts linked to the Chinese government, but went one step further when it elected to no longer accept paid state media ads that promote highly slanted coverage.[22] In 2020, as Hong Kong protests continued, both Twitter and Facebook introduced official account labeling in an attempt to enhance transparency and accountability.[23] Twitter reported that labels are intended to "provide additional context about accounts controlled […] by governments [and] state-affiliated media entities."[24] The company further specified that state-affiliated entities are only those that represent "the official voice of the nation state abroad" rather than individuals in their personal capacity.[25]

b) Assessing Strategy Effectiveness

Both Facebook and Twitter's strategies have found mixed success. Following Twitter's implementation of its labeling system in August 2020, Twitter users liked and shared significantly fewer tweets by CCP-controlled news outlets.[26] Nevertheless, tweets in non-English languages are not labeled to the extent that English language tweets are. Studies show that disclosure of state-run media is generally much weaker in languages other than English.[27] The language-based discrepancies vary by platform. On Facebook, 66% of all CCP accounts using the English language were labeled as of March 2021, whereas only 22% of CCP accounts in other languages were labeled at the time.[28] In contrast, 90% of all CCP accounts—regardless of language—were labeled

[21] *Removing Coordinated Inauthentic Behavior From China*, FACEBOOK NEWSROOM (Aug. 19, 2019), https://about.fb.com/news/2019/08/removing-cib-china/.
[22] *See* Vanessa Molter and Renee Diresta, *Pandemics & propaganda: How Chinese state media creates and propagates CCP coronavirus*, Harvard Kennedy School Misinformation Review (June 8, 2020), https://misinforeview.hks.harvard.edu/article/pandemics-propaganda-how-chinese-state-media-creates-and-propagates-ccp-coronavirus-narratives/.
[23] Schliebs, *supra* note 17, at 8.
[24] *Id.*
[25] *Id.*
[26] *Id.,* at 9.
[27] Molter and Diresta, *supra* note 22.
[28] Schliebs, *supra* note 17, at 9.

on Twitter.[29] The difference may be explained, in part, by Twitter's more aggressive approach to content takedowns. Historically, Facebook has taken a more permissive stance toward posts of government officials.[30]

More recent evidence suggests that Facebook has not gone far enough to mitigate misinformation in recent years.[31] As recently as October 2021, Francis Haugen, a former product manager for Facebook's civic integrity team,[32] testified before the Senate Commerce Committee that Facebook's efforts to combat misinformation are disingenuous.[33] Haugen has pointed to a particular Facebook algorithm that prioritizes user engagement over content authenticity.[34] Specifically, she suggested that content with the highest rates of engagement will be pushed to the top of Facebook's News Feed, regardless of its accuracy. Support for Haugen's accusations may be found in a 2019 internal report she copied. In the report, representatives from several European political parties revealed feeling pressure to take popular, albeit extreme, stances on Facebook in order to stand out on the platform.

Twitter's governance strategies may also leave something to be desired. A report by the Associated Press and the Oxford Internet Institute identified 26,879 accounts that managed to

---

[29] *Id*.

[30] Although the Chinese government's use of social media differed from that of President Trump and other world leaders who haven't shied away from representing themselves while using inflammatory language, the deceptive actions of the Chinese state are analogous in the sense they too are aimed to rouse tensions. *See* Miriam Berger and Elizabeth Dwoskin, *Trump ban by social media companies came after years of accommodation for world leaders who pushed the line*, WASHINGTON POST (Jan. 15, 2021), https://www.washingtonpost.com/world/2021/01/15/world-leaders-facebook-twitter-trump-ban/.

[31] *See e.g*, Elizabeth Dwoskin, *Misinformation on Facebook got six times more clicks than factual news during the 2020 election, study says*, WASHINGTON POST (Sep. 4, 2021), https://www.washingtonpost.com/technology/2021/09/03/facebook-misinformation-nyu-study/.

[32] Facebook's civic integrity team was charged with combatting election interference and misinformation, but was dissolved after the 2020 presidential election.

[33] *See* Alexandra S. Levine, et. al, *Whistleblower to Senate: Don't trust Facebook*, POLITICO (Oct. 5, 2021), https://www.politico.com/news/2021/10/05/facebook-whistleblower-testifies-congress-515083.

[34] *See* Jeff Horwitz, *The Facebook Whistleblower, Frances Haugen, Says She Wants to Fix the Company, Not Harm It,* WSJ (Oct. 3, 2021), https://www.wsj.com/articles/facebook-whistleblower-frances-haugen-says-she-wants-to-fix-the-company-not-harm-it-11633304122?mod=djemalertNEWS.

retweet Chinese diplomats or state-run media nearly 200,000 times before getting suspended, likely because takedowns came weeks and months after the state's misleading activity.[35]

c) Legal Implications

Twitter's and Facebook's efforts to curb state-sponsored misinformation about the Hong Kong protests raise important questions about freedom of expression on platforms that are increasingly considered public fora. There is no international consensus as to what qualifies as permissible speech and how such speech should be disseminated. Instead, some scholars suggest that "the determination of what appears online occurs through a process of contestation between three different groups forming the prongs of the free speech triangle—states, users, and technology companies."[36] What results, then, is a patchwork of rules from the public and private sectors that are subject to varied application across jurisdictions.

Moreover, the rules of tech companies' may clash with regional laws. In May 2021, the CCP issued a national security law that could force opposition groups to take down their social media accounts.[37] The law empowers police to demand that online service providers such as Facebook and Twitter release information about users and remove content the government deems a danger to national security.[38] Facebook and Twitter have expressed alarm over the security law, threatening not to comply with government requests for user data. However, the tech companies will likely be forced to choose between leaving Hong Kong altogether and complying with the new laws, since refusal to comply could result in large fines and months of imprisonment for local employees.

---

[35] Kintez, *supra* note 17.
[36] Hamilton, *supra* note 6.
[37] *See The impact of the national security law on Hong Kong*, REUTERS (May 31, 2021), https://www.reuters.com/world/asia-pacific/impact-national-security-law-hong-kong-2021-05-31/.
[38] *Id.*

2. *Case Study 2: Misinformation in the German QAnon Movement*

The proliferation of the QAnon movement in Germany demonstrates (1) the difficulty of reversing the proliferation of social media movements and (2) the interplay of public and private information governance regimes. QAnon is a conspiracy theory that emerged on the notoriously controversial message board 4chan when anonymous poster "Q Clearance Patriot" began posting cryptic messages in 2017.[39] The conspiracy reached its height during the pandemic, fueled by COVID-related misinformation.[40] Although the U.S. is the largest producer of QAnon-related content, the conspiracy has developed an international following, with the second highest amount of content coming from German-speaking Facebook users.[41] The German QAnon movement, not unlike its American counterpart, has gained support largely by appealing to fears about changing national demographics. Thus, some experts were not surprised that QAnon, which "has powerful echoes of the European anti-Semitism of centuries past," metastasized through German social media channels.[42]

a) Misinformation Mitigation Strategies

Although Facebook and Twitter strategically endeavored to inhibit QAnon-fueled misinformation, some say that the efforts came too late. In October 2020, Facebook announced plans to ban groups openly supporting QAnon conspiracy theories and restrict suggestions of QAnon content.[43] This change came on the heels of criticism that the tech company hadn't done

[39] *See* Kevin Roose, *YouTube Cracks Down on QAnon Conspiracy Theory, Citing Offline Violence*, N.Y. TIMES (update Sep. 29, 2021), https://www.nytimes.com/2020/10/15/technology/youtube-bans-qanon-violence.html.
[40] *See* Ciaran O'Connor, et. al, *The Boom Before the Ban: QAnon and Facebook*, ISD GLOBAL (2021), https://www.isdglobal.org/wp-content/uploads/2020/12/20201218-ISDG-NewsGuard-QAnon-and-Facebook.pdf.
[41] *Id.*
[42] *See* Katrin Bennhold, *QAnon Is Thriving in Germany. The Extreme Right Is Delighted*, N.Y. TIMES (Oct. 11, 2020), https://www.nytimes.com/2020/10/11/world/europe/qanon-is-thriving-in-germany-the-extreme-right-is-delighted.html.
[43] *See* Sheer Frenkel, *Facebook Amps Up Its Crackdown on QAnon,* N.Y. TIMES (Oct. 6, 2020), https://www.nytimes.com/2020/10/06/technology/facebook-qanon-crackdown.html.

enough in preceding months to curb the movement.[44] QAnon was likely able to skirt Facebook's earlier policies and retain its presence on the platform by renaming groups, toning down its language, and co-opting health and wellness groups about child safety.[45]

Twitter has also struggled to contain the QAnon movement on its platform, even after the FBI identified the group as a domestic terrorism threat.[46] Despite attempts to clamp down on QAnon in 2020,[47] by September of that year, the most popular QAnon Twitter accounts maintained a total of 2.4 million followers.[48] It wasn't until U.S. law enforcement agents identified QAnon supporters among those who stormed the Capitol on January 6 that Twitter found a suitable justification to delete 70,000 accounts "shar[ing] harmful QAnon-associated content at scale."[49]

b) Assessing Strategy Effectiveness

Experts have yet to reach consensus as to the best practices for mitigating the spread of QAnon conspiracy theories on social media. Some believe in a hard line approach to all conspiracy theories, while others recommend a tempered strategy that might include labeling conspiratorial content.[50] Hardliners point to sweeping bans like those of Reddit, formerly a breeding ground for QAnon theories, as strong models for eliminating misinformation.[51]

---

[44] Months earlier, Facebook instituted a policy in which it would bar only those accounts that explicitly incited violence. *See id.*
[45] *See id.*
[46] *See* Shirin Ghaffary, *Facebook and Twitter said they would crack down on QAnon, but the delusion seems unstoppable*, VOX (Oct. 6, 2020), https://www.vox.com/recode/21499485/qanon-facebook-twitter-bans-republican-politics.
[47] *See QAnon: Twitter bans accounts linked to conspiracy theory*, BBC (July 22, 2020), https://www.bbc.com/news/world-us-canada-53495316.
[48] *See* Ghaffary *supra* note 46.
[49] *An update following the riots in Washington, DC, TWITTER BLOG, @TwitterSafety (Jan. 12, 2021),* https://blog.twitter.com/en_us/topics/company/2021/protecting--the-conversation-following-the-riots-in-washington--.
[50] *See* Shira Ovide, *How Facebook Can Slow QAnon for Real*, N.Y. TIMES (Sep. 21, 2020), https://www.nytimes.com/2020/09/21/technology/facebook-qanon.html?action=click&module=RelatedLinks&pgtype=Article.
[51] *Id.*

Facebook's and Twitter's tactics to preclude QAnon misinformation on their platforms may have been more effective if they had coordinated their efforts with other social media companies from the start. Some scholars cite tech platforms' coordinated response to ISIS social media strategy as a model for mitigating extremist recruitment efforts online.[52] Once Facebook and Twitter began strategically tackling QAnon conspiracies, many of the movement's supporters fled to the encrypted messaging app Telegram. Unlike WhatsApp, which restricts group chat size to 256 users, Telegram groups can have up to 200,000 members, while public channels can have unlimited subscribers. At one point, one German Telegram channel for QAnon supporters had amassed 70,000 followers, spawning more than 1,200 local channels and groups.[53] Thus, while QAnon found its way overseas vis-a-vis the largest social media networks, it may maintain support through less mainstream platforms that are harder to police.

c) Legal Implications

Facebook's and Twitter's hesitations to crackdown on deceptive QAnon content earlier may reflect ineffective German legislation on harmful online speech. Despite Germany's strict laws on freedom of expression, the tech companies appeared to maintain adherence to more flexible American legal principles. In 2017, Germany passed the Network Enforcement Law ("NetzDG"), which requires social network operators to delete fake news and hate speech or face millions of euros in fines.[54] Nevertheless, some say that the law hasn't proven particularly effective in curbing misinformation.[55] This is likely because the law focuses primarily on explicitly

[52] *See* Ryan Broderick and Ellie Hall, *Tech Platforms Obliterated ISIS Online. They Could Use The Same Tools On White Nationalism*, BUZZFEED NEWS (March 20, 2019), https://www.buzzfeednews.com/article/ryanhatesthis/will-silicon-valley-treat-white-nationalism-as-terrorism.
[53] *See* Darren Loucaides, *How Germany became ground zero for the COVID infodemic*, OPEN DEMOCRACY (Mar. 31, 2021), https://www.opendemocracy.net/en/germany-ground-zero-covid-infodemic-russia-far-right/.
[54] *See* Daniel Leisegang, *No freedom to hate: Germany's new law against online incitement*, EUROZINE (Sep. 29, 2017), https://www.eurozine.com/no-freedom-to-hate-germanys-new-law-on-online-incitement/.
[55] *See* Janosch Delcker, *Germany's balancing act: Fighting online hate while protecting free speech,* POLITICO (Oct. 1, 2020), https://www.politico.eu/article/germany-hate-speech-internet-netzdg-controversial-legislation/.

defamatory and hateful speech, and sets high bars for qualifying as such.[56] Pushback that the law suppresses freedom of expression has also likely inhibited German courts from actually prosecuting potential violations.[57]

Given the reluctance of traditional governors of speech to punish the QAnon movement, social media companies may not be wholly blameworthy for their own hesitations. Private companies must balance content moderation with goals of preserving free speech, lest they receive backlash from users. Taking down content simply because it is not demonstrably true runs the risk of sanctioning widely accepted content such as religious beliefs.[58] Ultimately, law enforcement may need to take a more active role in linking online misinformation to real-world violence.[59] Once law enforcement agencies illuminate legitimate threats posed by organizations like QAnon, tech companies have far wider latitude to remove their content. When it comes to content moderation, social media companies may be more justified in playing a role akin to the judiciary: resolving imminent controversies for which there are clearly delineated public laws rather than issuing their own private legislation.

II. Conclusions

U.S. laws around content regulation are largely permissive due to the country's historical reverence for the First Amendment. Accordingly, it is difficult for U.S.-based social media platforms to adequately adapt their policies in foreign jurisdictions. Their efforts to do so raise questions of fairness and due process as they attempt to apply rules evenly across borders. Private companies have attempted to boost their legitimacy by both partnering with external organizations

[56] *Id.*
[57] *Id.*
[58] *See* Ghaffary *supra* note 46.
[59] For instance, once FBI investigations identified QAnon supporters among the January 6 Capitol rioters, Facebook and Twitter were able to justify sweeping takedowns of QAnon content. *See* TWITTER BLOG *supra* note 49.

and seeking community input. Yet, tech companies should take caution in encouraging digital forms of vigilantism. While it is possible that user input could lead to greater accountability in the fight against misinformation, it could also sow further discord and legal fragmentation. Given conflicting domestic standards for regulating speech, tech platforms should consult international human rights standards whenever possible. Tech companies should also be encouraged to recuse themselves from markets where states are the primary perpetrators of misinformation. Ultimately, the most successful tech-based strategies will likely be ones that contemplate public sector input and are flexible for later amendment.

Responding to Disinformation on Private Messaging Apps

I. Introduction

Disinformation is by now a well-known phenomenon that has global impact on not only political outcomes but also health outcomes as well, as illustrated by the worldwide surge in health-related misinformation during the covid-19 pandemic. However, the focus on disinformation spread has been mostly around social media and related networks. While that continues to be a serious issue, there has been less attention on the spread of disinformation through private messaging apps (PMA). Private messages have the potential to do great harm because they are afforded an extra layer of trust, as such messages are often exchanged directly through individuals with real-world relationships.

The potential magnitude of disinformation over PMA's is comparable to that on social media. While there are over 3.5 billion users of social media, such as Facebook; in 2021, an estimated 3.09 billion mobile phone users accessed private messaging apps to communicate.[1] Whatsapp is the largest messaging platform, with over 2 billion monthly active users worldwide.[2] For all the socially convenient and useful possibilities enabled by online platforms, there nonetheless remain significant risks to social cohesion, democracy, and public health when disinformation is widely shared. On private messaging apps, information can go viral in minutes as individuals forward messages along to their friends or groups, without any way to determine its origin. However, there has been relatively less scrutiny on the role of private messaging apps in the propagation of disinformation.

[1] *See* L. Cici, *Number of mobile phone messaging app users worldwide from 2018 to 2025*, STATISTA (Nov. 15, 2021) https://www.statista.com/statistics/483255/number-of-mobile-messaging-users-worldwide/
[2] Id.

II. Background

Private messaging applications (PMA) are installed on a mobile phone and allow users to send instant messages to other users over their phone's internet connection or mobile network. Messaging apps are capable of transmitting not only text-based messages and emojis, but can also feature voice and video calls, sending and receiving files, images, audio, and even location data. Illiteracy or lack of education is no barrier for PMA users; a graphical interface assists these users to communicate effectively through audio, video, and images. Although these messaging apps were initially designed for private communication between individuals or small groups, they are now increasingly being used as mass messaging apps, with forming "shadow social networks" capable of broadcasting messages and other media to thousands of people at once.[3] Thus, messages containing misinformation have a capacity to spread virally over PMAs through direct forwarding to large groups, even though the information is not accessible to the general public as on open forum, such as Facebook, Twitter, and YouTube.[4]

In contrast to these open forums, many messaging apps are semi-anonymous and offer end-to-end encryption of messages to further preserve the privacy of users, such that only the participants can see the content of the messages unless it is specifically flagged and forwarded to service providers. While end-to-end encryption provides a heightened level of privacy, it makes content moderation more challenging and makes tracking the spread of messages, as well as the dissemination of disinformation, nearly impossible from the developer side. However, some

[3] Brian X. Chen & Kevin Roose, *Are Private Messaging Apps the Next Misinformation Hotspot*, New York Times (Feb. 3, 2021) https://www.nytimes.com/2021/02/03/technology/personaltech/telegram-signal-misinformation.html Group chats on Whatsapp are now limited to 250 users while Telegram has a size limit of 200,000 users.

[4] *See* Tony Romm, *Fake cures and other coronavirus conspiracy theories are flooding WhatsApp, leaving governments and users with a 'sense of panic,'* The Washington Post (Mar. 2, 2020) https://www.washingtonpost.com/technology/2020/03/02/whatsapp-coronavirus-misinformation/

platforms do collect metadata related to users and messages, such as where, when, and how the app is used.[5]

Further compounding the issue of disinformation spread is the cross flow and reinforcement loop of information between public platforms and PMAs. Actors can proliferate disinformation by hosting content on multiple open platforms, such as YouTube and Twitter, which is subsequently shared to groups on private messaging apps; simultaneously, private message apps are utilized to reinforce disinformation by sharing such content to drive additional traffic to content on open platforms.[6] This cross flow of information between private and public platforms raises questions about the appropriate way to counter malicious actors on public platforms. For example, deplatforming an individual or group spreading misinformation on a public forum may cause that individual or group to pivot onto a private forum or directly spread disinformation through PMAs, where it is more difficult for authorities and researchers alike to track their activities. Recipients of disinformation through these private channels are subsequently free to share and disseminate those messages on open platforms, allowing malicious actors to continue their influence campaigns.[7]

Most significantly, because the intimacy of private messaging apps can contribute to perceptions of content credibility, a major concern about the spread of false information on private messaging apps is not necessarily virality, but the level of trust placed in the content that is shared. Lack of access and erosion of trust in mainstream news media creates a trust vacuum that misinformation on private messaging apps can occupy.[8] Furthermore, information shared between

---

[5] *See WhatsApp Privacy Policy*, WhatsApp, https://www.whatsapp.com/legal/privacy-policy?eea=0 (last accessed Mar. 24, 2022)
[6] *See* REUTERS, Reuters Institute Digital News Report 2018 at 6 (2018).
[7] *See* Liz Harrington (@realLizUSA), Twitter, https://twitter.com/realLizUSA?ref_src=twsrc%5Egoogle%7Ctwcamp%5Eserp%7Ctwgr%5Eauthor
[8] *See* Elizabeth Dwoskin & Annie Gowen, *On WhatsApp Fake News is Fast – and can be fatal*, The Washington Post (Jul. 23, 2018) https://www.washingtonpost.com/business/economy/on-whatsapp-fake-news-is-fast--and-can-

contacts in private messaging apps is often uniquely tailored to specific recipients and encourages sincere engagement with content spread privately.[9] A survey conducted by the American Press Institute and the AP-NORC Center found that the identity of the information sharer impacted the level of trust the recipient felt toward the information.[10] Information received through a private messaging app is seen as more credible, or is more likely to be read, when it is received directly from a contact known to the person in real life, rather than open social media platforms. A separate research study conducted in 2018 found that news content received on WhatsApp was more likely to be trusted than news found on Facebook, which was largely due to the personal nature of the app, such as the close relationship with the sender and the directness of content dissemination.[11]

III. Policy Challenges in Regulating Private Messaging Apps

There are inherent tensions between mitigating online harms and competing democratic values of free expression and privacy. Content moderation of digital communications can easily infringe on free expression to an undesirable degree; while a lack of content moderation can risk the further spread of harmful content, which posing real-world consequences. Furthermore, there are significant risks to privacy rights and civil liberties when private messaging app providers are regulated. The core dynamic of the encryption debate is the tension between preserving privacy and mitigating online harms. End-to-end encryption is seen as an important technical bulwark

---

be-fatal/2018/07/23/a2dd7112-8ebf-11e8-bcd5-9d911c784c38_story.html ("In many countries, messaging services are the main platform to get online," said Samantha Bradshaw, co-author of the report from the Computational Propaganda Project. "The closed platforms can be more dangerous because the information is spreading in these intimate groups of friends and family — people we tend to trust.")

[9] WhatsApp does not see itself as a social-media service, because content is not posted publicly and algorithms do not spread information virally.

[10] *See Who Shared It? How Americans Decide What News to Trust on Social Media*, Associated Press NORC, https://apnorc.org/projects/who-shared-it-how-americans-decide-what-news-to-trust-on-social-media/ (last accessed Mar. 24, 2022)

[11] *See* Antonis Kalogeropoulos, *The Rise of Messaging Apps for News*, *in* Reuters Institute Digital News Report 2018, https://s3-eu-west-1.amazonaws.com/media.digitalnewsreport.org/wp-content/uploads/2018/06/digital-news-report-2018.pdf

against threats to individual rights, democracy, and national security. However, others criticize the constraining effect that encryption has on law enforcement and security agencies.

IV. Case Studies

A. *India*[12]

The social and legal struggles in India with disinformation spread through WhatsApp provide a lesson in how over regulation of PMAs can spill over to affect fundamental constitutional freedoms of expression and the press and a right to privacy. From 2016, WhatsApp disinformation resulted in over 24 deaths by lynch mobs in various rural villages. Furthermore, the Indian elections of 2019 featured widespread use of WhatsApp disinformation perpetrated by politicians and volunteers recruited by political parties to stoke racial tensions between Hindu and Muslim communities to mobilize and energize the party's political base.

Despite initial resistance, WhatsApp fully cooperated with Indian authorities in response to WhatsApp fueled violence and introduced points of friction to slow the spread of disinformation and disrupt the strong trust afforded to such messages received. WhatsApp developers limited the number of people who could participate in group chats as well as the number of times a message could be forwarded in a single instance. Additionally, message recipients could see whether a message was one that was forwarded many times. These methods successfully hindered the spread of misinformation.[13] WhatsApp also offered media literacy training and ran full-page newspaper ads.[14]

---

[12] India has a literacy rate of 77.7 % and 21.5% of the population aged 25 and above hold a bachelor's degree or higher. Trust in media 38%, 2021 (average for region); *see Education at a Glance 2018*, OECD, 42-46 (Sep. 11, 2018)

[13] *See* Melo, Philipe & Vieira, Carolina & Garimella, Kiran & Vaz de Melo, Pedro & Benevenuto, Fabrício, *Can WhatsApp Counter Misinformation by Limiting Message Forwarding?* (2019) https://www.researchgate.net/publication/335926199_Can_WhatsApp_Counter_Misinformation_by_Limiting _Message_Forwarding

[14] Rishi Iyengar, *WhatsApp is using newspapers to fight fake news in India*, CNN (Jul. 10, 2018) https://money.cnn.com/2018/07/10/technology/whatsapp-india-newspaper-ads-fake-news/index.html

During this time, the Indian government also amended the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules under the Information Technology Act. This revision required a "significant social media intermediary" that primarily provide messaging services, such as WhatsApp, and Facebook Messenger to establish an office physically based in India, and appoint a locally residing "Chief Compliance Officer" and "Chief Grievance Offer;" authorize government official to require tech companies to take down social media posts it deemed unlawful; to track the origin of messages, which essentially requires developers to break end-to-end encryption to comply; monthly compliance report disclosing details of complaints received and action taken, as also details of contents removed pro-actively; and intermediaries that failed to comply with the new rules could have its immunity against liability and criminal prosecution revoked.[15]

In response, WhatsApp filed a lawsuit against the Government of India in the Delhi High Court, alleging that the new rules violate the right to privacy of Indian users, and calling for the intermediary rules to be declared unconstitutional.[16] Other companies, including mainstream media companies in India, filed similar suits.

B. *Taiwan*[17]

[15] *See The Information Technology (Intermediary Guideleins and Digital Media Ethics Code) Rules, 2021,* Indian Ministry of Electronics and Information Technology, §139(E) Part II-III (Feb. 25, 2021)

[16] *See* Mike Isaac, *WhatsApp Sues India's Government to Stop New Internet Rules*, The New York Times (May 21, 2021) https://www.nytimes.com/2021/05/25/technology/whatsapp-india-lawsuit.html

[17] Taiwan has a literacy rate of 98.70% and 46.5% of Taiwan's population aged 15 and above hold a bachelor's degree or higher. However, it consistently has one of the lowest scores for trust in media in the region; Taiwanese tend not to trust in news media because the line between news and entertainment is very blurred as a result of circular reporting and competitive pressures within the mass media industry to be first to report a story. Rumors that often originates from reinforced by social media and private message rumor mills are reinforced and reinvigorated by news media coverage of such rumors. Additionally, the independence of news media a Financial Times report quoted journalists at China Times and CtiTV, both owned by a pro-Beijing business tycoon, as saying they receive daily calls from Beijing to shape news coverage. *See Education*, Government Portal of the Republic of China (Taiwan) https://www.taiwan.gov.tw/content_9.php (last accessed Mar. 24, 2022); *2020-2021 Taiwan at a Glance*, Ministry of Foreign Affairs, Republic of China (Taiwan) 24-30 (2021); Emily Feng, *Taiwan Gets Tough on Disinformation Suspected from China Ahead of Elections*, NPR (Dec. 6, 2019)

Taiwan tops the list of countries that receive the most disinformation spread by a foreign government with significant Chinese influence in media and disinformation campaigns across both traditional and social media.[18] Despite these challenges, Taiwan has successfully minimized the impact of such disinformation, during both its 2020 presidential election and the Covid-19 pandemic.[19]

LINE is the most popular private messaging app in Taiwan with 21 million monthly active users, in a country with a population of just 23.8 million people. The app features an integrated news platform and private and encrypted group chats. Lines actively discourages sharing content to other apps by requiring extra steps for external sharing. Additionally, 47% of respondents surveyed for the Digital News Report revealed that they use LINE for news.[20] However, according to a LINE survey, 46% of respondents revealed that they had seen or received suspicious content in app. However, only 33% of respondents revealed fact-checked these messages, while only 25% had shared accurate information with others.

Taiwan's success in combating disinformation campaigns on PMA's was ultimately due to successful cooperation between the government, civil society, and social media platforms and legal measures taken by the Taiwanese government. Tsai's administration fostered public-private

https://www.npr.org/2019/12/06/784191852/taiwan-gets-tough-on-disinformation-suspected-from-china-ahead-of-elections

[18] *See* Valeria Mechkova et. al., *Measuring Internet Politics: Introducing the Digital Society Project*, Digital Society Project (May 2019); *Democratic Taiwan Battling Disinformation From China Ahead of Elections*, Radio Free Asia Nov. 6, 2018) https://www.rfa.org/english/news/china/democratic-taiwan-battling-disinformation-11062018111310.html

[19] Examples of the harmful effects of misinformation in Taiwan includes a diplomat's suicide following a widely circulated, but falsified story about rescue efforts to evacuate Taiwanese tourists stranded in Japan in the aftermath of Typhoon Jebi. In the weeks leading up to Taiwan's 2020 presidential election, there was a surge of disinformation spread through LINE that attempted to influence voters or suppress turnout, such as false claims of ballot tampering, electoral fraud, and even a SARS outbreak that would endanger voters. Nevertheless, Chinese disinformation did not have a significant effect to boost performance of their preferred candidates at the polls; Tsai won by a large majority.

[20] *See Taiwan*, Digital News Report, https://www.digitalnewsreport.org/survey/2020/taiwan-2020/ (last accessed Mar. 24, 2022)

partnerships which integrated civic technology developed by private companies and developers. Through its "g0v" initiative, the government supported small teams of developers to devise technological solutions to civic issues, including disinformation; for example, this initiative supported the development of CoFacts, a collaborative fact checking database and chatbot, which was eventually integrated within a broader LINE Fact Checking program. Additionally, instead of seeking to clarify disinformation through traditional press releases or resorting to takedown orders and censorship, the administration took active anti-trolling measures, using memes and humorous messages to counter disinformation.[21] The administration required each ministry to respond rapidly to disinformation with such counter measures, which were often widely shared and reached the press and the public even before disinformation did. Additionally, the government created education initiatives to promote information and media literacy, for both children and elderly populations.[22] For example, media literacy trucks, including a Facebook mobile classroom, were deployed across the country to conduct fake news identification workshops for citizens with less media experience in the weeks before the 2020 presidential election.[23] The administration also worked directly with providers, including Facebook and Line to reduce disinformation spread. LINE provided a "rumors clarification" section free of charge to the Taiwanese government on its main news page in addition to its fact checking program.

The government also has legal recourse to deter and prosecute individuals who spread disinformation through private messaging apps with laws including the Social Order Maintenance Act, and the recently promulgated Anti-Infiltration Law, Communicable Disease Control Act, and

[21] Given Taiwan's recently history of marital law, such measures have been staunchly avoided by subsequent administrations to avoid comparison to past repressive regimes and policies. Taiwan Turns to Facebook and Viral Memes to Counter China's Disinformation. *See* Chun Han Wong & Phillip Wen, *Taiwan Turns to Facebook and Viral Memes to Counter China's Disinformation*, Wall Street Journal (Jan. 3, 2020)
[22] https://mlearn.moe.gov.tw/
[23] LINE also provides free resources for K-12 educators.

Special Act for Prevention, Relief and Revitalization Measures for Severe Pneumonia with Novel Pathogens. Although arrests under SOMA have increased in recent years, with 151 in 2019 and 233 in the first five months of 2020 alone, the majority of SOMA cases do not lead to convictions, nor significant penalties like prison terms or steep fines. Users of private messaging apps have been arrested and fined for posting covid-19 related disinformation. However, legal action is further complicated by challenges of attribution as well as a lack of distinction between mis- and dis-information spreading. Additionally, the prolonged length of time involved in such cases do not allow for prompt response and resolution of spreading disinformation.

V. Policy Proposals

An effective disinformation inoculation strategy for private messaging apps will require close cooperation between government and private entities to devise an approach that can strike a reasonable balance between privacy, free expression, and security. To that end, the U.S. must take an active role in shaping tech policy while encouraging and guiding private message app developers to adopt practices to increase friction in order to reduce disinformation spread over their apps because most of the world's most frequently downloaded and utilized PMAs are founded, headquartered, and developed in the States.[24]

As seen in the successful example of LINE's integration of CoFacts in Taiwan, U.S. agencies can similarly foster the development of fact checking groups, mechanisms, and bots in addition to promulgating minimum standards for such efforts. Specifically, the Federal Trade Commission, Federal Communications Commission, and National Telecommunications and Information Administration within the Department of Commerce are the three U.S. agencies have

[24] *See* Mark Williams, *Secure Messaging Apps Comparison*, https://www.securemessagingapps.com (last visited Dec. 10, 2021).

varying degrees of jurisdiction and regulatory authority through both formal and informal mechanisms to provide guidance to PMA developers and companies. Mechanisms, which include guidance on optimal group chat size limits, simultaneous forwarding limits, and integrated bots can then be incorporated into apps while preserving user privacy with no need to break encryption. Individual messages can be forwarded directly to fact checking bots or accounts without casting message app providers as "arbiters of truth." Moreover, user-initiated fact checking through integrated mechanisms would provide real-time feedback and encourage users to continuously engage with the platform, which is beneficial for both service providers and users.[25] Most importantly, the U.S. government should clarify transparency obligations to further study the spread and impact of disinformation over PMAs. Regulations to promote greater transparency would require reporting by service providers that details internal policies toward addressing disinformation, the number and content of disinformation related complaints, and whether and how metadata on users is collected and utilized. This would provide more information for developers, researchers, and policymakers to better understand and prevent the spread of disinformation going forward. Specific regulations and recommendations to introduce or increase friction to slow the spread of disinformation are among the least intrusive to the constitutionally protected freedoms of expression and press. Furthermore, regulators, instead of stringent rules, can offer informal guidance through workshops and white papers in support of PMA providers to further avoid potential infringement on free speech protections.

Furthermore, as disinformation is a global issue, various programs and organizations under the aegis of the United Nations (UN) should coordinate a new initiative or add to the mission of

[25] *See* Elizabeth Lange & Doowan Lee, *How One Social Media App Is Beating Disinformation*, Foreign Policy (Nov. 23, 2020), https://foreignpolicy.com/2020/11/23/line-taiwan-disinformation-social-media-public-private-united-states.

an existing initiative to outline frameworks and best practices for governments and private companies under a whole-of-society approach that prioritizes development, human rights, transparency, and education to prevent and minimize the harmful effects of disinformation. The UN could establish a new program that involves officials representing the following offices – UN Global Pulse, the Office of the Secretary-General's Envoy on Technology, the United Nations Department of Global Communications, United Nations Development Programme (UNDP), United Nations Educational, Scientific and Cultural Organization (UNESCO), World Health Organization (WHO), and Office of the United Nations High Commissioner for Human Rights (OHCHR). Each of these offices has previously been involved with disinformation inoculation policies from various policy lenses. For example, UNESO has undertaken the development of comprehensive strategies for media and information literacy campaigns and has experience working with the UN Global Pulse initiative as well as the WHO to develop strategies to counter Covid-19 related misinformation.[26] Participation of existing UN development groups can also ensure policies and regulations are sensitive to the needs and challenges of less developed states. Furthermore, the involvement of the OHCHR is essential to ensure that further proposals do not infringe on human rights, particularly as one-on-one communication, such as messages facilitated through private messaging apps, falls under the ambit of Article 17 of the International Covenant on Civil and Political Rights and Article 8 of the European Convention on Human Rights that establish and protect a right to digital privacy. By combining the expertise and experience that various UN agencies and organizations already possess, this multi-lateral group would effectively identify and coordinate specific strategies and aid to promote social cohesion, stability, and education as a long-term approach to prevent the harms associated with misinformation

[26] *See UN entities partner to strengthen work to counter the COVID-19 infodemic*, UN Global Pulse (Sept. 11, 2020) https://www.unglobalpulse.org/2020/09/un-entities-partner-to-strengthen-work-to-counter-the-covid-19-infodemic/.

Addressing Disinformation Using International Law: Existing Instruments and Challenges

I. Introduction

This chapter explores the possibility of using international law to combat disinformation. It reviews the applicability of existing international legal instruments to combat disinformation, and the general challenges of applying international law to such a task. It finds that, while many of these international legal instruments could be construed to address acute issues related to disinformation, international law lacks comprehensive legal tools for combatting disinformation. Beyond those gaps exist further challenges to relying on existing international law to address disinformation.

II. International Legal Instruments for Combatting Disinformation

International law lacks comprehensive legal tools for combatting disinformation. A number of existing international instruments could be interpreted in order to combat disinformation. These include: the 1936 International Convention on the Use of Broadcasting in the Cause of Peace; the human right to self-determination; the concept of State sovereignty; and Article 8 of the International Law Commission's Articles on the Responsibility of States for Internationally Wrongful Acts (ARSIWA). While each of these could be construed to apply to disinformation, they leave behind gaps in international law.

A. 1936 International Convention on the Use of Broadcasting in the Cause of Peace

The treaty is not the most helpful for dealing with disinformation, because it only applies to false news, even though distorted news can be just as harmful.

Article 3(1) states that, "the high contracting parties mutually undertake to prohibit and, if occasion arises, to stop without delay within their respective territories any transmission likely to harm good international understanding by statements the incorrectness of which is or ought to be known to the persons responsible for the broadcast."[1] The English-language of the treaty ("broadcasting") is broad enough to include mass media, while the French-language ("radiodiffusion," denoting the means of broadcasting rather than content), might not be.[2] But, according to the Vienna Convention on the Law of Treaties (VCLT), conflicting language versions of a treaty should be resolved according to the treaty's object and purpose.[3] The treaty was ratified in response to budding technologies that, for the first time, allowed information to reach people directly, while territorial States were unable to counter it.[4] As such, the treaty likely applies to mass media like television and the internet. Moreover, for the treaty to apply, there must be (1) a transmission of information; (2) that the person responsible for the broadcast knows or ought to know is incorrect; and (3) that is likely to harm "good international understanding."

In practice, the treaty's application is tricky, because it only applies to false news, even though distorted news can be just as harmful. The 2016 "Lisa case" provides an example of this. In early 2016, a thirteen-year-old Russian-German girl named Lisa disappeared in Berlin.[5] While accurately reporting the facts, Russian media, including *Sputnik* and *RT*, framed the disappearance as highlighting the security issues caused by the ongoing refugee crisis in Germany.[6] Russian

[1] International Convention on the Use of Broadcasting in the Cause of Peace art. 3(1), Sept. 23, 1936, 186 L.N.T.S. 301.
[2] Björnstjern Baade, *Fake News and International Law*, 29 EUR. J. INT'L. L. 1357, 1368 (2018).
[3] *Id.*
[4] *Id.* at 1365.
[5] *Id.* at 1359.
[6] *Id.* at 1360.

Foreign Minister Sergey Lavrov echoed the narrative during a press conference, while German Foreign Minister Frank-Walter Steinmeier rejected the narrative as "political propaganda."[7]

Although this was a (1) transmission of information that (3) likely harmed "good international understanding," at least between Ministers Lavrov and Steinmeier, Article 3(1) likely did not apply to this scenario, because no incorrect statements were made. Further, because *Sputnik* and *RT* both operate in Berlin, broadcasts from those locations would not fall under the "respective territories" of Russia, meaning they would not be covered by the treaty.

The treaty does not serve as a comprehensive solution to disinformation, because it does not cover distorted information, which can be just as harmful as false information.

B. The Right to Self-Determination

The international legal right to self-determination is likely not a comprehensive basis for combatting disinformation.

Under customary international law, and many different international agreements, people have the right to decide their own destiny in the international order.[8] For example, self-determination is protected in the United Nations Charter and the International Covenant on Civil and Political Rights as a right of "all people." In liberal democracies particularly, the election process is the final expression of a people's sovereign will.

According to declassified U.S. intelligence reports, Russia engaged in extensive disinformation campaigns during the U.S. election of 2020 to damage the candidacy of Joe Biden

[7] *Id.* at 1360.
[8] *Self determination (international law)*, CORNELL LAW SCHOOL LEGAL INFORMATION INSTITUTE, https://www.law.cornell.edu/wex/self_determination_(international_law) (last visited Sept.30, 2021).

and boost that of Donald Trump.[9] This followed similar activities during the election of 2016.[10] It can be argued that by attempting to substitute the will of the people with that of the Russian State, Russia likely violated the self-determination of the American people.[11] However, this argument is unlikely to hold water in an international court.

First, self-determination has typically been invoked in the context of determining the legitimacy of a new State.[12] Whether international lawyers are willing to expand the principle of self-determination beyond such contexts is unclear. Second, invoking self-determination in the context of disinformation campaigns assumes that the will of the American people can be determined before an election.[13] However, research has shown that, but for Russia's disinformation campaign, the election of 2016 might have turned out differently.[14]

The international legal right to self-determination is likely not a comprehensive basis for combatting disinformation.

C. State Sovereignty

The international principle of State sovereignty is likely not a comprehensive basis for combatting disinformation.

It is a basic principle of international law that because a State has the exclusive power to regulate the internal affairs of its "domaine réservé," one State will not interfere with the

[9] Julian E. Barnes, *Russian Interference in 2020 Included Influencing Trump Associates, Report Says*, N.Y. TIMES (May 27, 2021), https://www.nytimes.com/2021/03/16/us/politics/election-interference-russia-2020-assessment.html.
[10] S. Rep. No. 116-290 (2020).
[11] Jens David Ohlin, *Did Russian Cyber Interference in the 2016 Election Violate International Law*, 95 TEX. L. REV. 1579, 1596 (2018).
[12] *Id.*
[13] *Id.*
[14] Jane Mayer, *How Russia Helped Swing the Election for Trump*, NEW YORKER (Sept. 24, 2018), https://www.newyorker.com/magazine/2018/10/01/how-russia-helped-to-swing-the-election-for-trump.

sovereignty of another State.[15] The UN General Assembly's Friendly Relations Declaration, which reflects custom, clarifies that intervention is not limited to the use of physical force.[16]

The Tallinn Manual 2.0 discusses whether disinformation campaigns, particularly election meddling, constitute violations of sovereignty. Although it does not create new law, the document facilitated by the NATO Cooperative Cyber Defence Center of Excellence clarifies over 150 rules of international law.[17] Rule 66 on "Intervention by States" prohibits coercive intervention, which is interference that includes a coercive act with the potential to compel the target State to "engage in an action that it would otherwise not take."[18] Thus, in order for disinformation to violate State sovereignty, the disinformation must then function to eliminate a State's opportunity to engage in certain actions that it otherwise would.

State sovereignty likely does not serve as a comprehensive solution to disinformation because it only protects against intervention, even though interference can be harmful, as well. In the case of election meddling, a State's opportunities to act freely are not eliminated unless, for example, the opportunity to vote is removed because a candidate is killed or election infrastructure is destroyed.[19]

The international principle of state sovereignty is likely not a comprehensive basis for combatting disinformation.

[15] Ohlin, *supra* note 11, at 1596.
[16] *Id.* at 1587. (citing G.A. Res. 2625 (XXV), (Oct. 24 1970); Lori Fisler Damrosch, *Politics across Borders: Nonintervention and Nonforcible Influence over Domestic Affairs*, 83 AM. J. INT'L L. 1, 5–6 (1989)).
[17] *Id.* at 131.
[18] *Id.*
[19] *Id.*

D. Article 8 of the International Law Commission's Articles on the Responsibility of States for Internationally Wrongful Acts (ARSIWA)

In many cases, it may be helpful to combat disinformation by legally attributing it to a certain State. But because such attribution is difficult to establish, ARSIWA is not a comprehensive tool for combatting disinformation.

Privately conducted disinformation can be attributed to a State under Article 8 of the International Law Commission's Articles on the Responsibility of States for International Wrongful Acts (ARSIWA), which has become customary international law, if it can be shown that (1) the State directed or controlled the activity or (2) the State instructed the disinformation.[20] Further, the International Court of Justice's holding in the *Nicaragua* case suggests that heavy funding by a State in contribution to privately conducted acts of disinformation is insufficient to attribute the activity to the State.[21]

In practice, this means that many private acts of disinformation that are supported by sovereign States likely cannot be attributed to them. For example, the activity of news organizations like *Sputnik* and *RT* cannot necessarily be attributed to Russia, even though they are heavily funded by the State.[22] Even conduct publicly accepted as a State's own might not be attributable to the State in an international court. For example, although over 300 journalists were awarded medals by Russian President Vladimir Putin for their favorable coverage of the Crimean conflict, their acts of disinformation are not attributable to the State unless they continue thereafter.

[20] Baade, *supra* note 2, at 1361.
[21] *Id.* at 1362.
[22] *Id.* at 1361–1362.

As such, ARSIWA likely is also not a comprehensive tool for combatting disinformation. While many of these international instruments address acute issues related to disinformation, they likely do not serve to comprehensively address the issue.

III. General Challenges to Applying International Law

Beyond the challenge of construing existing international law to combat disinformation exist additional hurdles to effectively addressing disinformation. First, while international law can address false news, it leaves behind a gap in which distorted news continues. Additionally, the Westphalian underpinnings of international law make it difficult to arrive at objective truths in the international arena, which are arguably essential to combatting disinformation. Finally, assuming a claim under international law overcomes these challenges, it will face the slow pace of international action.

A. False v. Distorted Information

False news presents incorrect information, while distorted news is more subtle. In the latter, "information can be framed and presented in such a way as to make its recipients likely to draw certain (false) conclusions," even if the information itself is accurate.[23] Oftentimes, truthful news is distorted in order to incite hate or sow discord.[24] Although there are sizable limitations, as with the Broadcasting Convention discussed above, false news can be regulated under international law. However, regulation of distorted news is much more difficult. The *Lisa Case* highlights a gap in international law that allows harmful disinformation campaigns, so long as they do not disseminate false information.

[23] *Id.* at 1359.
[24] *Exposing Russia's Effort to Sow Discord Online: The Internet Research Agency and Advertisements*, U.S. HOUSE OF REPRESENTATIVES, PERMANENT SELECT COMMITTEE ON INTELLIGENCE, https://intelligence.house.gov/social-media-content/ (last visited Sept. 30, 2021).

B. State-Sovereignty

State sovereignty reigns supreme in international law. Every State has sovereignty over its territory and nationals, which entitles it to regulate the existence, conduct, and relationships of the various people, places, and things that relate to its territory, its nationals, or the State. This poses incredible challenges to any international effort to regulate disinformation, for two primary reasons:

(1) First, because every State has such sovereignty, they are only bound by international laws to which they consent, either by becoming party to a treaty or by acquiescing to an international instrument and helping it become customary international law.

(2) Second, in the context of disinformation, it likely is not possible to construct an objective determination of truth, which is necessary to regulate information, in such a State-centric, sovereignty-forward conception of international law. States can choose to believe what is in their best interests.[25]

C. The Slow Pace of International Action

Any claim under international law will have to overcome the slow pace of international action. For example, a case before the International Court of Justice takes a notoriously long amount of time to resolve. In contrast, technological weaponry can be deployed instantly. This makes it difficult rely on international law to address issues of disinformation in a timely fashion

[25] The simplest example of this is Russia's consistent dismissal of criticism over its election interference, confirmed by the U.S. intelligence community. Russia not only denied such allegations, but even justified the hacks as warranted. Sophie Tatum, *Putin denies election attack but justifies DNC hack because 'it is true'*, CNN (July 16, 2018), https://www.cnn.com/2018/07/16/politics/putin-fox-interview/index.html.

and suggests that such issues could be better addressed by domestic systems, such as legislative or military/intelligence bodies, or by the private sector.[26]

IV. Conclusion

While many of these international legal instruments could be construed to address acute issues related to disinformation, many gaps remain, beyond which exist further challenges to relying on international law to address disinformation.

[26] In the United Kingdom, for example, Ofcom is authorized under the Communications Act of 2003 ("the Act") to set standards for broadcasting content. Rule 5.5 of the Act states that: "Due impartiality on matters of political or industrial controversy and matters relating to current public policy must be preserved on the part of any person providing a service [...]. This may be achieved within a programme or over a series of programmes taken as a whole". OFCOM BROADCAST BULLETIN, Issue 288, 20 (Sept. 21, 2015). In 2015, pursuant to the Act, Ofcom found that RT violated portions of the act by not showing "due impartiality" and by materially misleading the audience. Specifically, RT did not preserve due impartiality with matters of political controversy in Ukraine, presenting "a negative picture of the Ukrainian Government and its military forces," including allegations of rape and murder. *Id.* at 55. Although Ofcom presents an example of effective domestic regulation of disinformation, the potential to abuse its regulatory powers is concerning because all news organizations face the challenge of presenting objective and factual reporting.

Geopolitical Competition

As a result of the reach and ease of use of online platforms and social media, disinformation campaigns conducted on the internet are becoming the frontline of geopolitical competition. Disinformation campaigns have evolved significantly from Cold War-era propaganda campaigns into a multitude of nuanced strategies tailored to the goals and context of the involved states. Disinformation is becoming an increasingly effective manner of directly influencing foreign and international affairs, from manipulating foreign economies and trade negotiations to sowing dissent and distrust in foreign governments and interfering with their electoral processes.

This Section explores the characteristics of disinformation campaigns originating from some of the largest and most involved disinformation actors including Russia, China, and the United States. The ways in which these countries manipulate and disseminate disinformation against foreign states provides insight into the underlying geopolitical goals and strategies of these countries, as well as the ways in which technological evolutions have facilitated state-to-state disinformation campaigns.

The case studies in this Section demonstrate how disinformation campaigns are adapting to evolutions in technology and changing political fields. These changing strategies have been accompanied by parallel changes in the tools and techniques that states are using in combating foreign disinformation campaigns to varying levels of success. As this Section demonstrates, the variability of disinformation campaigns and rapid technological advancements that facilitate disinformation require similarly nuanced approaches to combatting disinformation that are tailored to specific states' strategies.

Disinformation Fuels a Shadow Trade War:

China's Use of Disinformation as Both a Justification and a Tool for Economic Retaliation

I. Background

In April 2020, Australia called for an investigation into the origins of COVID-19, including an inquiry into China's management of the initial outbreak in Wuhan.[1] In apparent response, China began to increase tariffs on key Australian exports. Two-way trade with China is worth 11.7 percent of Australia's gross domestic product (GDP), and key industries include wine, beef, coal, wheat, and services.[2] This disagreement was not the first that China and Australia have had over the years,[3] however China has this time relied on disinformation as both a justification and a tool for retaliation on Australian goods. Though China has attempted to either justify or conceal its economic retaliation on Australian goods, its actions violate the terms of the WTO agreement to which it, and Australia, are parties. This paper will first discuss how China has used trade measures to retaliate for Australian disinformation though this is a violation of the WTO agreement. It proceeds to discuss how China has invoked its own use of disinformation to conceal its retaliatory reasons for imposing restrictions on Australian goods, also in violation of the WTO agreement.

[1] Brett Worthington, *Marise Payne calls for global inquiry into China's handling of the coronavirus outbreak*, ABC NEWS (Apr. 19, 2020, 2:48 AM), https://www.abc.net.au/news/2020-04-19/payne-calls-for-inquiry-china-handling-of-coronavirus-covid-19/12162968.

[2] James Laurenceson & Michael Zhou, *Understanding Australia's Economic Dependence on China*, AUSTRALIA-CHINA RELATIONS INSTITUTE (June 21, 2019), https://www.australiachinarelations.org/content/understanding-australias-economic-dependence-china.

[3] China has previously responded with tariffs for action taken by Australia to ban Chinese tech giants, Huawei and ZTE from providing 5G technology to Australia. Lee Jeong-ho, Kristin Huang & Darwin, *Australia at Centre of Strategic Tussle between US and China*, SOUTH CHINA MORNING POST (Aug. 3, 2019, 12:51 PM), https://www.scmp.com/news/china/diplomacy/article/3021294/darwin-australia-centre-strategic-tussle-between-us-and-china.

This paper will not discuss alternative WTO agreements and how they may apply, such as the Technical Barriers to Trade or Sanitary and Phytosanitary Measures agreements, or China's Accession Protocol.

II. China's Accusation of Australian Disinformation Spurs Economic Retaliation

China has reacted to Australian calls for an investigation into the origins of COVID-19 with accusations that Australia is spreading disinformation about China and its handling of the outbreak. In retaliation, China has enacted higher tariffs on several key Australian exports. Economic retaliation for political circumstances is not permitted under the WTO agreement to which both Australia and China are signatories.[4]

A. China's Accusation of Australian Disinformation

China's response to Australia's initiation of a COVID-19 origin investigation began with discreditation of Australia through direct accusations by officials.[5] Such statements focused predominantly on Australia's relationship with the United States in attempts to influence Australian domestic understanding of their government's motivations for inquiring into the COVID-19 origins.[6] The Chinese embassy in Australia stated that Australian politicians were "cooperat[ing] with the U.S. in its propaganda war against China" which "exposes the former's ignorance and bigotry as well as their lack of independence in serving orders from others."[7] In addition, Ambassador Cheng Jingye, the Chinese ambassador to Australia personally stated that

[4] WTO Agreement: Marrakesh Agreement Establishing the World Trade Organization, Apr. 15, 1994, 1867 U.N.T.S. 154, 33 I.L.M. 1144 (1994) [hereinafter WTO Agreement].
[5] Ashley Townshend, *China's Pandemic-Fueled Standoff with Australia*, WAR ON THE ROCKS (May 20, 2020), https://warontherocks.com/2020/05/chinas-pandemic-fueled-standoff-with-australia/.
[6] *Id.*
[7] Interview by *Global Times* with Chinese Embassy Spokesperson (Ap. 21, 2020).

Australia was "pandering to… Washington" and making a "political move to please a certain country," in attempts to further influence Australian domestic sentiments toward the investigation.[8]

China also claims that Australia's call to investigate COVID-19 origins is itself disinformation that China must correct. China blatantly dismisses the comments of Marise Payne, the Australian Minister for Foreign Affairs, who discussed Australia's reasons for wanting a COVID-19 investigation, as "inconsistent with the facts."[9] Not only is China attempting to use disinformation itself as a mechanism for fomenting unrest within the Australian populace regarding the COVID-19 investigation, but even more critically it is asserting that Australia is spreading disinformation and following that assertion with economic restrictions.

In conjunction with its discreditation of Australia for its initiation of an investigation into the origins of COVID-19, China almost immediately threatened economic harm to Australia through Chinese disengagement from key trade industries between the countries. In a discussion about the Australian initiation of a COVID-19 investigation, Ambassador Cheng warned that Australia's supposed disinformation tactic will make Chinese tourists and students "think 'Why should we go to such a country that is not so friendly to China?' and ordinary people might think 'Why should we drink Australian wine? Eat Australian beef?'"[10] Within weeks the threat of consumer boycotts became a reality, and China had banned beef from four Australian exporters and announced an 80% tariff on barley.[11] The threats made by Ambassador Cheng were unique because they were overt, calibrated to be feasible, and relied on a novel approach of combining economic pressure and a disinformation campaign against Australia.[12] China reacted to what it

[8] Interview by Andrew Tillet with Cheng Jingye, Chinese Amassador to Australia (Apr. 27, 2020).
[9] Geng Shuang, Chinese Foreign Ministry Spokesperson, Regular Press Conference (Apr. 20, 2020).
[10] Interview by Andrew Tillet with Cheng Jingye, *supra* note 7.
[11] Su-Lin Tan, *China's Restrictions on Australian Beef, Barley Seen as Retaliation for Support of Coronavirus Investigation*, SOUTH CHINA MORNING POST (May 12, 2020, 8:53 PM), https://www.scmp.com/economy/global-economy/article/3084062/chinas-restrictions-australian-beef-barley-seen-retaliation.
[12] Townshend, *supra* note 4.

claimed was Australian disinformation regarding COVID-19 by threatening retaliation on key Australian industries.

Though the Ambassador's threat of harm to Australian industry and the Chinese government's trade-restrictions on the same industries indicate that the imposed restrictions were a direct reaction to the initiation of the Australian investigation into the origins of COVID-19, the Chinese government initially denied such assertions. The Chinese Foreign Ministry alleged that the barley tariff was a "normal trade remedy investigation."[13] The *Global Times*, an official Chinese state-owned media outlet, ran a series of articles denying the increased tariff rates were related to the Australian COVID-19 investigation while warning Australia that it will have "much bigger problems than barley if it continues to take unfriendly action."[14] However, the Chinese Foreign Ministry spokesman Zhao Lijian dropped this pretense in July 2021, stating that China "will not allow any country to reap benefits from doing business with China while groundlessly accusing and smearing China and undermining China's core interests based on ideology."[15] Blatant economic retaliation for political issues has aroused concerns about Chinese violation of the WTO agreement.

Though the Chinese government claims that the trade restrictions were imposed in compliance with international trade law as a part of normal trade remedy investigations and requirements, Australia appears to question the legitimacy of this argument, bringing lawsuits to the WTO insinuating that China's protectionist actions are merely a retaliation for Australia's COVID-19 investigation.

[13] Zhao Lijian, Chinese Foreign Ministry Spokesperson, Regular Press Conference (May 11, 2020).
[14] Wang Bozun, *Tariffs on barley not the only problem Australia may face*, GLOBAL TIMES (May 10, 2020), https://www.globaltimes.cn/content/1187941.shtml.
[15] Zhao Lijian, Chinese Foreign Ministry Spokesperson, Regular Press Conference (July 6, 2020).

B. Australia-China Legal Action Before the WTO

There are currently three cases pending before the World Trade Organization (WTO) between China and Australia, two of which were filed by Australia, and one that was filed (seemingly in response) by China. The WTO's predecessor, the General Agreement on Tariffs and Trade (GATT) was an agreement created to pursue peace and security, avoid terms of trade losses, and establish a commitment against domestic interests. The GATT has been fully incorporated into the WTO, which binds all member states to act in concordance with the agreement. China and Australia are both members of the WTO and are thus bound to comply with the WTO agreement, which has made all forms of trade protectionism other than domestic subsidies and tariffs illegal and prevents countries from raising tariff rates above an agreed upon bound rate. There are some exceptions to the WTO objective of reducing trade tariffs, such as anti-dumping and countervailing duty tariffs, but even those may not be applied out of retaliation for another country's political statements.[16] Only once countries have won an appellate decision on their case (if it is appealed), may they take retaliatory trade action against the other member country, unless the other member country withdraws the protectionist measures at issue.

The first case currently pending before the WTO between China and Australia is a complaint by Australia related to the tariffs the Chinese imposed on barley under anti-dumping and countervailing duty grounds. Consultations at the WTO were requested on 16 December 2020, and a WTO panel was established on 28 May 2021. In their Request for Consultation, Australia

[16] The only exceptions for surpassing the bound tariff rate are countries' rights to impose value added taxes, fees, or other charges commensurate with services rendered, and anti-dumping and countervailing duties. Anti-dumping duties are enacted on imported products when they are priced below fair market value, while countervailing duties are levied on imported goods that are subsidized by foreign governments often allowing them to sell the product at a lower price. Should member states assert that another member has violated the WTO, the two members enter an arbitration phase which, if unsuccessful, results in proceeding to a panel of judges to decide on the issue. The member states may, after the panel decision, decide to appeal the case to the WTO appellate body, though it is currently inactive.

advanced many arguments about why the anti-dumping and countervailing duty orders were legally inadequate, including Chinese methodology, interpretation of terms, and failure to meet procedural requirements.[17] Critically, Australia advanced an argument related to China's imposition of the tariffs as a retaliation for what China saw as Australian disinformation about the origins of COVID-19. Though Australia did not specifically discuss Chinese retaliation, it alleged that China "improperly initiated investigations on the basis of applications that were not made "by or on behalf of the domestic industry" and that "China failed to determine, on the basis of an examination of the degree of support for, or opposition to" said anti-dumping and countervailing duty measures.[18] Anti-dumping and countervailing duty measures can only be brought by governments when initiated by or on behalf of the domestic industry and Australia appears to allege that there was no industry support for these duties but the Chinese government has imposed them of their own determination as retaliation.[19]

China's bound tariff rate for barley seed not for sowing is 3 percent and the bound tariff rate for barley for sowing is 0 percent.[20] This means that China would be unable to raise the tariff rates on those products to the level of 80.5 percent like they did using regular tariff adjustment or

[17] Request for Consultations by Australia, *China – Anti-dumping and Countervailing Duty Measures on Barley from Australia*, WTO Doc. WT/DS598/1 (Dec. 21, 2020).
[18] *Id*. Australia cited Chinese violations of Articles 5.1,5.2 and 5.4 of the Anti-Dumping Agreement and Articles 11.1, 11.2 and 11.4 of the Subsidies and Countervailing Measures (SCM) Agreement, both actionable before the WTO.
[19] Marrakesh Agreement Establishing the World Trade Organization, Apr. 15, 1994, 1867 U.N.T.S. 154 [hereinafter Uruagay Round Agreement], Agreement on Implementation of Article VI of the General Agreement on Tariffs and Trade 1994 [hereinafter "Anti-Dumping Agreement"], Article 5.4. The agreement states that "an application shall be considered to have been made 'by or on behalf of the domestic industry' if it is supported by those domestic producers whose collective output constitutes more than 50 per cent of the total production of the like product produced by that portion of the domestic industry expressing either support for or opposition to the application. However, no investigation shall be initiated when domestic producers expressly supporting the application account for less than 25 per cent of total production of the like product produced by the domestic industry." Request for Consultations by Australia, *China – Anti-dumping and Countervailing Duty Measures on Barley from Australia*, WTO Doc. WT/DS598/1 (Dec. 21, 2020).
[20] Tariff Data, World Trade Organization, http://tariffdata.wto.org/TariffList.aspx.

they would violate the WTO.[21] However, the anti-dumping and countervailing duty exception provides for governments to go beyond their bound WTO tariff rates in order to remedy a harm to their domestic industry.[22] Anti-dumping and countervailing duty tariffs may not be applied out of retaliation for another country's political statements. They may only be applied due to concern from a significant portion of domestic industry about harms to their production, and not from the government alone or for reasons beyond the impact on domestic production.[23]

While the WTO has not yet had a panel sitting on the issue, it seems Australia will argue before the panel that China was in fact unable to apply the anti-dumping and countervailing duty exceptions to the bound tariff rate as the tariffs were driven by the government rather than the domestic industry itself. Given Chinese officials' threats to particular industries in response to the Australian investigation and the close timing of the imposition of the threatened tariffs a few weeks later, as well as Mr. Zhao's statements that China deliberately retaliated for the COVID-19 investigation, Australia will likely make the case that this shows ulterior motives for the imposition of anti-dumping and countervailing duties other than industry requests. If Australia can prove this, then the WTO panel should find China in violation of the treaty. However, given the lack of transparency within the Chinese government, it may be a difficult argument for Australia to win on at the WTO, particularly given close connections between Chinese government and industry. This case poses a critical moment where the WTO can either show that countries cannot conceal political trade retribution as domestic industry harm or show that a government can act retributively without violating the WTO agreement merely by attributing the action to their

[21] Tariff Data, World Trade Organization, http://tariffdata.wto.org/TariffList.aspx:; Jens Kastner & Dimitri Simes, *China tariff on Australia's barley reshapes global trade*, NIKKEI ASIA (Mar. 19, 2021), https://asia.nikkei.com/Business/Agriculture/China-tariff-on-Australia-s-barley-reshapes-global-trade.
[22] GATT art. VI.
[23] *Id*.

domestic industry.

Though China cannot legally base its imposition of tariffs on accusations of disinformation by the Australian government, should the WTO approve of Chinese action in these cases it will grant governments leeway to justify the imposition of tariffs for political reasons, such as accusations of disinformation spreading, by claiming a convenient harm to their domestic industry. The legal standard regarding this issue is clear: tariffs surpassing a country's bound rate justified by political reasons and not subject to a WTO exception are not legal.[24] However, the lack of informational transparency and the inherent difficulty of drawing clear connections between government trade action and unlawful motivations make it difficult for a WTO panel to decide when the laws have been violated.

III. China's Use of Disinformation to Justify Trade Restrictions on Australian Goods

China has decided to employ disinformation as a tool to conceal the true reason for its economic retaliation against Australia. China used disinformation to explain their orders to domestic industry not to purchase Australian products, in one instance citing environmental standards as its reasoning for the ban on Australian coal, while increasing imports of coal from other countries. China's use of disinformation to conceal the true reason for its government direction to domestic industry not to purchase goods from one country is a violation of the WTO agreement.

A. China's Direction to Domestic Industry

Under GATT Article XI, General Elimination of Quantitative Restrictions, members to the

[24] GATT art. VI & art. II.

WTO cannot place "prohibitions or restrictions other than duties, taxes or other charges, whether made effective through quotas, import or export licenses or other measures" against any other member state. This ban against any trade restriction other than tariff measures applies only to government conduct, not private parties.[25] Thus, actions taken by private parties to restrict trade can only be challenged at the WTO if government action was essential to the restriction.[26]

Though no WTO decisions technically have precedential value, the oft-cited case that establishes governments can be held responsible for controlling the actions of their private sector, is Japan-Semiconductors.[27] In 1986, an arrangement was struck between Japan and the U.S. to improve U.S. market access in Japan and avoid Japanese dumping. The agreement required the government of Japan to request that Japanese producers and exporters of semi-conductors not export semi-conductors to the U.S. at prices below company-specific costs. The European Community challenged the arrangement as a violation of Article XI. In particular, Japan argued that none of their actions were legally binding and that Japanese companies were not technically restricted in any way, stating that the exports were being limited by private enterprises for their own self-interests.[28] However, Article XI refers to "measures," which the WTO panel interpreted as including actions that were not legally binding, as long as there was significant incentive or disincentive for the producers and exporters to perform in a certain way and government interference was essential to that action. The Panel concluded that the government of Japan was operating "to exert maximum possible pressure on the private sector" and "essentially dependent on government action."[29] Thus, the agreement violated Article XI.

[25] Panel Report, *Japan – Trade in Semi-Conductors*, L/6309 – 35S/116 (adopted May 4, 1988).
[26] *Id*.
[27] *Id*.
[28] *Id*.
[29] *Id*.

On the other hand, the WTO has ruled that certain situations involve government action that is too attenuated from the private conduct that they do not qualify as Article XI border restrictive measures. In the Argentina-Bovine Hides case, Argentina instituted a regulation which required members of the Argentinian leather industry to be present during the customs clearance process for the export of bovine hides.[30] The European Communities argued that the presence of members of the leather industry was an attempt by the Argentinian government to use peer pressure to restrict exports.[31] However, the WTO ruled it was too difficult to attribute any trade disruption to the presence of a member of the leather industry when they did not have actual power to halt bovine exports.[32] The WTO does not regulate private conduct, and the government action in this case was too attenuated to violate Article XI.[33]

China has taken government action to warn certain domestic industries not to purchase Australian products. One such industry that the Chinese government verbally told key domestic importers to no longer purchase from Australia, is coal.[34] In October 2020, China warned state-owned utilities and steel mills to stop importing Australian thermal and coking coal with immediate effect.[35] China's verbal warning to key domestic producers not to import certain materials, though it carries no legal binding just like Japan's warning to exporters, would likely be seen as sufficient government interference. This is even more likely to be found by the WTO given that the companies that China warned not to purchase Australian coal were state-owned enterprises

[30] Panel Report, *Argentina – Measures Affecting the Export of Bovine Hides and the Import of Finished Leather*, WT/DS155 (adopted Feb. 16, 2001).
[31] *Id*.
[32] *Id*.
[33] *Id*.
[34] Su-Lin Tan, *China-Australia relations: coal shipments continue but remain stuck off Chinese coast amid ban*, SOUTH CHINA MORNING POST (May 13, 2021, 4:30 AM), https://www.scmp.com/economy/article/3133225/china-australia-relations-coal-shipments-continue-remain-stuck-chinese?module=inline&pgtype=article.
[35] Jenny Ma & Jessie Li, *Chinese state-owned end-users given verbal notice to stop importing Australian coal*, S&P GLOBAL (Oct. 9, 2020), https://www.spglobal.com/platts/en/market-insights/latest-news/coal/100920-chinese-state-owned-end-users-given-verbal-notice-to-stop-importing-australian-coal-sources.

that are even more directly influenced by the government than private corporations would be and were therefore sufficiently "disincentivized" from buying Australian coal by nature of their close relationship with the government.

B. Chinese Disinformation as a Rationalization

China's attempt to justify their verbal warning to producers not to purchase Australian coal came a month later disguised as "environmental quality" concerns.[36] This appears to be a form of disinformation in an attempt to manipulate the narrative surrounding China's ban on Australian goods for political retaliation, as the increased inspections on coal for environmental justifications have only had significant impact on Australian imports, leaving $700 million sitting idle off the Chinese coast.[37] This makes for a suspicious justification in the wake of the verbal ban given to Chinese producers and mounting Chinese-Australian tensions, indicating this is a form of Chinese disinformation used to justify producer restrictions on Australian goods. Under the WTO, countries can impose quality inspections on imported goods as part of their regulation standards.[38] However, it must not violate Articles I or XIII which establish the MFN principle prohibiting different treatment of countries.[39] If Australia faces an import ban on a product that other countries are still permitted to export to China, this constitutes a de facto violation of the WTO agreement under the most-favored-nation principle discussed in Articles I and XIII.[40] This is what has occurred with China's restriction of the importation of Australian coal, as Australia has been

[36] Bill Birtles, *China claims 'quality' problem with Australian coal as $700 million worth sits idle off ports*, ABC NEWS (Nov. 25, 2020), https://www.abc.net.au/news/2020-11-26/china-claims-quality-problem-with-australian-coal/12921354.
[37] *Id*.
[38] GATT art. VIII.
[39] GATT art. I & art. XIII.
[40] AB Report, *Canada – Certain Measures Affecting the Automotive Industry*, WT/DS139/AB/R (adopted June 19, 2000).

unable to export coal to China while Mongolia has increased its exports to China to fill the market that Australia lost.[41]

Though China attempted to justify the ban that it asked producers to impose on Australian coal with environmental quality concerns, this approach merely amounted to spreading disinformation about the reasons for Chinese economic action without making it any more defensible under WTO law. While coal is only one example in which China warned domestic producers not to import certain products from China, it clearly demonstrates China's willingness to impose retaliatory restrictions on Australian products and then rely on disinformation to justify its trade protectionism.

C. Application of Chinese Domestic Trade Law

While China violated its obligations under the WTO agreement, it may not have violated its domestic trade laws. In the Foreign Trade Law of the People's Republic of China, Article VI, it states that China will "on the principle of reciprocity grant the other party most-favored-nation treatment."[42] Though China makes the same commitment to providing equal treatment to other countries as is required under the MFN principle in the WTO agreement, it includes a broader list of reasons that can justify state intervention restricting or prohibiting the import or export of goods

[41] One exception under which China could argue the regulation and ban on Australian coal could be justified is Article XX(g): "relating to the conservation of exhaustible natural resources." Since China cites environmental concerns for their regulation, and mining coal contributes to pollution and the erosion of natural resources, at first glance this seems a possible exception that China could rely on. However, the exception also stipulates that the "measures are made effective in conjunction with restrictions on domestic production or consumption." Given that the restriction on Australian coal led to larger importations of coal from Mongolia and greater domestic production, China's ban did nothing to prevent more coal from being produced and thus China will not be able to claim this exception. China may instead be able to argue under article XX(b): "necessary to protect human, animal or plant life or health." This exception would likely require a demonstration by China that the importation of coal negatively impacts human, animal or plant life or health in China. This is a similarly difficult exception for China to meet given that the ban on Australian coal merely shifted China's supply to other markets but did not reduce it, which contributes equally to the detriment of human, animal or plant life or health in China, thus making the measure unjustifiable.

[42] Foreign Trade Law of the People's Republic of China, 2004.

from other countries.[43]

China, in banning the import of Australian coal, likely could not justify that action under any of the WTO exceptions, but it may be able to satisfy one of the many exceptions listed in Article XVI of its Foreign Trade Law. Some of the possible exceptions that could justify China's actions include: "(7) the import needs to be restricted in order to establish or accelerate the establishment of a particular industry" or "(9) the import needs to be restricted to maintain the State's international financial status and the balance of international payment."[44] While China did not invoke either of these justifications for its actions, given the breadth of these exceptions and the discretion they allot to the judgement of the government, it seems likely that their ban on Australian coal may be more easily justified under domestic law compared to international law. This may be important because even if Australia were to bring this case to the WTO and win, given the current state of the WTO, China would be able to appeal the case into the void, and thus prevent Australia from being able to legally enact retaliatory measures. Should that occur, one of Australia's only other opportunities for remedy would be through Chinese domestic courts applying Chinese law. Additionally, should Australian coal producers choose to sue, their only option for remedy would be to sue in Chinese courts.[45]

IV. Conclusion

China has relied on disinformation as its justification for trade protectionist measures against Australia. On the one hand, China accused Australia of spreading disinformation and thus

[43] *Id*.
[44] *Id*.
[45] Australian industry cannot sue in Australian courts as the only action Australian courts could take would be to impose countervailing duties, antidumping duties, or safeguards, and the requirements for those measures to apply under the WTO require there to be imports causing the domestic industry harm, which is not what is occurring here.

retaliated with increased tariffs and restrictions, attempting to conduct a shadow trade war for political reasons. On the other hand, China has resorted to using disinformation to publicly justify their protectionist measures imposed on Australian goods to conceal the political motives behind their heightened trade protections. Both relying on accusations of Australian disinformation and using Chinese disinformation to conceal political motives as an excuse to increase tariffs violate WTO law and likely Chinese law as well.

China's Deliberate Use of State Disinformation to Justify National Security Measures: Case Studies of Xinjiang and Hong Kong

I. Introduction

In the last three decades, China has emerged on the world stage as one of the most powerful nations, with an economy that is expected to overtake that of the United States by 2028.[46] China also ranks first among the countries with the most internet users.[47] While observers have expected such developments to come with greater political openness, political power is centralized and firmly rooted in the Chinese Communist Party under a system of "one-party rule."[48] As China's social fabric has developed in response to greater wealth and technological advancement, China's leaders have prioritized countering threats to the party's control over opening to the outside world.[49] This paper will explore how China has engaged in deliberate disinformation campaigns to counter such threats and, in doing so, constructed its own narrative to justify its controversial policies. Part I will discuss of how China disseminates disinformation. Part II will analyze two disinformation case studies in Hong Kong and Xinjiang to demonstrate how the Chinese government employs disinformation campaigns to rally support around repressive national security measures. Part III will conclude with a discussion of the legality of these measures under international legal principles.

[46] *Chinese economy to overtake US 'by 2028' due to Covid*, BBC (Dec. 26, 2020), https://www.bbc.com/news/world-asia-china-55454146.
[47] *See* Joseph Johnson, *Countries with the highest number of internet users as of Q1 2021*, STATISTA (July 19, 2021), https://www.statista.com/statistics/262966/number-of-internet-users-in-selected-countries/.
[48] Michael Bogdan, *The International Legal Status of Governing Political Parties in One-Party States*, 22 GERMAN Y.B. INT'L L. 335 (1979).
[49] Jacques deLisle & Avery Goldstein, *To Get Rich Is Glorious: Challenges Facing China's Economic Reform and Opening at Forty* 9 BROOKINGS (2019) https://www.brookings.edu/wp-content/uploads/2019/04/9780815737254_ch1.pdf.

II. Background and Censorship

In order to contextualize China's use of state-sponsored disinformation campaigns, it is necessary to analyze its system of governance as a starting point. The Chinese Communist Party has had a monopoly on state power since the establishment of the People's Republic of China in 1949.[50] Though there have been periods of economic reform that foreign observers hoped would yield positive changes in civil society, the state has consistently suppressed freedom of speech, most notably in the Tiananmen Square Massacre of 1991 where the state cracked down on student-led pro-democracy demonstrations.

The increase in Chinese internet users and emergence of new communication platforms has transformed civil society and online discourse. From 2008 to 2020, the number of Chinese internet users increased nearly three-fold to 989 million users.[51] China's mobile application use has also skyrocketed; in 2019 Chinese users accounted for one third of global consumer spending on mobile applications.[52] In response to these seismic social shifts, the State has had to employ new tactics to control the public narrative. These changes have occurred amidst Xi Jinping's consolidation of power, which has resulted in domestic tightening and the suppression of ideological, ethnic, and linguistic differences.[53]

III. China's Use of State Disinformation

In recent years, the CCP has sought to impose its Chinese world view and increase its "discourse power" both domestically and internationally through the deliberate dissemination of

[50] Orville Schell, *Life of the Party: How Secure Is the CCP?*, 100 FOREIGN AFF. 68 (2021).
[51] *Id*.
[52] Lai Lin Thomala, *Mobile apps in China - statistics & facts*, STATISTA (June 18, 2021), https://www.statista.com/topics/5577/mobile-apps-in-china/.
[53] Brandon Zheng, *Centenary Propaganda and Nationalism with Xi Jinping Characteristics for a New Era*, BAKER INSTITUTE FOR PUBLIC POLICY, https://www.bakerinstitute.org/files/17553/.

information. Discourse power, which is "the concept that a country can attain increased geopolitical power by setting agendas internationally through influencing the political order and values both domestically and in foreign countries," is often realized through targeted information operations.[54] China's use of such information operations to achieve soft power objectives is not in isolation; in the past five years governments across the globe have been using social media to influence public opinion and investing resources in propagating disinformation.[55] A study drawing upon data from the Computational Propaganda Project's investigation into global social media manipulation reveals that, in authoritarian regimes, the entities disseminating propaganda are often government ministries.[56] As an authoritarian regime, China's dissemination of information is state-sponsored and highly formalized. This section will analyze how China domestically conducts state disinformation campaigns to promote stability while internationally it conducts such campaigns to sow instability.

A. State Media

Since the founding of the PRC, official state media outlets have been a key part of China's attempt to control and influence the public narrative. In recent years, Chinese media outlets have become a flashpoint in ongoing tensions between the U.S. and China. In 2020, the United States designated nine Chinese state media outlets as "foreign missions" controlled by Beijing.[57] David Stilwell, the assistant secretary for East Asia and the Pacific at the State Department, said that

[54] *Chinese Discourse Power: China's Use of Information Manipulation in Regional and Global Competition*, ATLANTIC COUNCIL (December 2020), https://www.atlanticcouncil.org/wp-content/uploads/2020/12/China-Discouse-Power-FINAL.pdf.
[55] *Id*.
[56] Samantha Bradshaw & Philip Howard, *The Global Organization of Social Media Disinformation Campaigns*, 71 J. OF INT'L AFFAIRS 23 (2018).
[57] Robert Delaney, *US designates 5 Chinese state media outlets as Beijing operatives*, SOUTH CHINA MORNING POST (Feb. 19, 2020), https://www.scmp.com/news/world/united-states-canada/article/3051246/us-imposes-restrictions-5-chinese-state-media.

these entities "are not independent news organizations" and are "effectively controlled" by the Chinese Communist Party.[58]

The international and domestic impact of disinformation disseminated by Chinese state media can be seen in the context of the recent COVID-19 epidemic. As the pandemic spread, China used its media structure to promote positive narratives about China's handling of the coronavirus to promote domestic stability.[59] State media has also cast doubt on the origins on the virus.[60] When a Chinese Foreign Ministry spokesman made a groundless statement that COVID originated in a lab in Fort Detrick, MD, CCP publication Global Times started an online petition, which gained 25 million signatures, to investigate the lab's facilities.[61]

B. Social Media

Social media has been a powerful tool for the state to disseminate disinformation in China. The process of state disinformation on social media platforms is both top down and bottom up.[62] China's "wolf warriors" are state officials who spread disinformation through social media accounts.[63] The "keyboard army," also known as netizens or the "50 Cent Army" is comprised of millions of citizens who monitor the internet and influence public opinion. The Chinese government has been suspected of hiring nearly two million people to post on social media, as if

---

[58] Edward Wong, *U.S. Designates Four More Chinese News Organizations as Foreign Missions*, N.Y. TIMES (June 22, 2020), https://www.nytimes.com/2020/06/22/us/politics/us-china-news-organizations.html.
[59] Julia Bergin, Johan Lidberg & Louisa Lim, *The COVID-19 Story: Unmasking China's Global Strategy*, INTERNATIONAL FEDERATION OF JOURNALISTS (May 2021), https://www.ifj.org/fileadmin/user_upload/210512_IFJ_The_Covid_Story_Report_-_FINAL.pdf.
[60] Joshua Kurlantzick, *How China Ramped Up Disinformation Efforts During the Pandemic*, COUNCIL ON FOREIGN RELATIONS (Sep. 10, 2020), https://www.cfr.org/in-brief/how-china-ramped-disinformation-efforts-during-pandemic.
[61] Austin Ramzy & Amy Chang Chien*, Rejecting Covid Inquiry, China Peddles Conspiracy Theories Blaming the U.S.*, N.Y. TIMES (Oct. 12, 2021) https://www.nytimes.com/2021/08/25/world/asia/china-coronavirus-covid-conspiracy-theory.html.
[62] Renee Diresta & Vanessa Molter, *Pandemics & propaganda: How Chinese state media creates and propagates CCP coronavirus narratives*, HARVARD KENNEDY SCHOOL MISINFORMATION REV. (June 8, 2020), https://misinforeview.hks.harvard.edu/article/pandemics-propaganda-how-chinese-state-media-creates-and-propagates-ccp-coronavirus-narratives/.
[63] *Id*.

they were representing their genuine opinion, to advocate the government's position.[64] These groups not only to spread propaganda, but also to distract the public through shifting the narrative away from contentious issues.[65]

Returning to the pandemic, we can see the State disinformation apparatus at work through social media channels. At the beginning of 2020, in response to the lack of information regarding the emergence of COVID-19, there were calls on Chinese social media platforms for freedom of speech and more transparency for whistleblowers.[66] After censoring these types of posts, the Chinese government, with assistance from the 50-Cent Army undertook a campaign to question the origins of the virus, while touting the successes of the Chinese government in stopping the spread and engaging in diplomatic efforts to help other countries struggling with COVID-19.[67] Simultaneously, China engaged in an effort to sow panic about the virus in the United States, through social media posts and text message campaigns.[68] This shows broad reach of Chinese social media and how the government has strategically used these channels both as a stabilizing force domestically and destabilizing force internationally.

III. Case Studies: China's Disinformation Campaigns in Hong Kong and Xinjiang

From justifying crackdowns in Tibet to fostering disunity in Taiwan, there are many examples of how the Chinese government has fostered disinformation domestically and abroad to promote its official narrative. This Part will focus on two issues: China's use of disinformation to

[64] *Id*.
[65] Gary King, et al*., How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, not Engaged Argument*, 111 AM. POL. SCI. REV. 484 (2017).
[66] Raj Verma, *China's diplomacy and changing the COVID-19 narrative*, 75 INT'L J. 248 (2020).
[67] *Id.*
[68] Joshua Kurlantzick, *How China Ramped Up Disinformation Efforts During the Pandemic*, COUNCIL ON FOREIGN RELATIONS (Sep. 10, 2020), https://www.cfr.org/in-brief/how-china-ramped-disinformation-efforts-during-pandemic.

justify the enactment of the Hong Kong National Security Law in June 2020 and China's campaign to control the narrative surrounding human rights abuses in Xinjiang.

A. Hong Kong's National Security Law

In recent years, China has been using its disinformation apparatus to undermine pro-democracy protesters both domestically and abroad. As part of the handover agreement of Hong Kong to China in 1997, Hong Kong's common law institutions and traditions were preserved under the "one country, two systems" principle, which ensured citizens with freedoms of assembly, speech, and press.[69] Hong Kongers have frequently exercised these rights, particularly in the mass protest movements of 2003 over a proposed national security law, in 2014 to advocate for a democratic selection process of Hong Kong's chief executive, and in 2019 to oppose an extradition bill allow extraditions from Hong Kong to the mainland.

In 2019, China created an alternate narrative which portrayed the protesters as a small, violent gang, backed by foreign agents that were trying to create division and strife within mainland China.[70] Through painting the protesters as American-backed terrorists, the Chinese government stoked nationalistic sentiment on Chinese social media platforms. Though most of the protests were peaceful, the violent scenes served to undermine the movement's pro-democracy goals.[71]

While the 2019 extradition law was eventually retracted by Hong Kong leadership, the false narratives surrounding the protests served a larger purpose for the Chinese Government. In June 2020, Hong Kong's Chief Executive promulgated a new national security law, *The Law of*

[69] Roderick Munday, *Hong Kong's Final Court of Appeal: The Development of the Law of China's Hong Kong*, 74 CAMBRIDGE L.J. 155 (2015).
[70] Steven Lee Myers & Paul Mozur, *China is Waging a Disinformation War Against Protestors*, N.Y. TIMES (August 13, 2019), https://www.nytimes.com/2019/08/13/world/asia/hong-kong-protests-china.html?.
[71] *Id.*

the People's Republic of China on Safeguarding National Security in the Hong Kong Special Administrative Region to respond to "strong obstruction and interference from anti-China forces disrupting Hong Kong and external hostile forces."[72] This law has been exceedingly controversial both within Hong Kong and across the globe, due to the its far-reaching extraterritorial scope and the wide range of conduct that it criminalizes.

The PRC and Hong Kong governments justified the National Security Law as a necessary measure to stabilize Hong Kong, after a year of violent protests and turmoil.[73] National People's Congress Chairman Wang Chen stated that the protesters "incited Hong Kong people to be anti-China and anti-Communist Party, … deliberately undermined social order in Hong Kong," and "violently confronted police enforcing the law."[74] The National Security Law was therefore justified by a narrative that was propagated by state disinformation.

B. Xinjiang

In April 2017, the Chinese Community Party, with the stated intention to eradicate domestic terrorism, began a "re-education" campaign in Xinjiang province in Western China.[75] As part of this campaign, ethnic minorities, including Uyghurs and Muslims, have been incarcerated in forced labor camps. Today it is estimated that as many as a million individuals have been detained in 300 to 400 facilities where they are subject to political indoctrination.[76] Despite

[72] Susan V. Lawrence & Michael F. Martin, Cong. Rsch. Serv., R46473, China's National Security Law for Hong Kong: Issues for Congress, at 4 (2020).
[73] *Id.*
[74] *Wang Chen Gives Explanation on 'Draft Decision of NPC on Establishment of Sound Legal System, Implementation Mechanism for Safeguarding of National Security in Hong Kong Special Administrative Region'*, XINHUA (May 22, 2020), http://www.xinhuanet.com/politics/2020-05/22/. c_1126019468.htm.
[75] Emma Iannini, *Cultivating Civilization: The Confucian Principles behind the Chinese Communist Party's Mass Imprisonment of Ethnic Minorities in Xinjiang and What Human Rights Advocates Can Do to Stop It*, 53 N.Y.U. J. INT'L L. & POL. 189 (2020).
[76] Beth Van Schaack & Maya Wang, et al., *"Break Their Lineage, Break Their Roots:" China's Crimes against Humanity Targeting Uyghurs and Other Turkic Muslims*, HUMAN RIGHTS WATCH & MILLS LEGAL CLINIC OF

overwhelming evidence of human rights violations in the region, the Chinese government has continued to deny these accusations.[77] An analysis of China's disinformation campaign reveals how the Chinese government is able to justify this narrative because it has created an alternative set of facts that it purports to be the truth.

There are two main facets to China's disinformation campaign in Xinjiang. First, it has distracted from these allegations through portraying positive narratives about the policies in the regions.[78] The Chinese government and state-owned media outlets have funded and amplified a YouTube channel which has videos depicting Xinjiang as idyllic place where ethnic groups live in harmony and in peace.[79] Similarly, China has produced a musical to show the ethnic harmony in Xinjiang, entitled "The Wings of Song."[80] The Chinese government has also invited journalists on government-run tours which paint a rosy picture of the situation in the camps and tout the successes of their efforts in combatting extremism.[81]

Second, China has deliberately injected disinformation into the domestic and global discourse. This tactic has been growing in popularity since early 2020, as state media and the Chinese government have been using social media to push alternate narratives. For example, it has used Facebook and Twitter to allege that the statistical data on Xinjiang is misleading, that Western media organizations reporting on Xinjiang are biased in their research, and that the policies in the

---

STANFORD LAW SCHOOL, April 19, 2021 https://law.stanford.edu/publications/break-their-lineage-break-their-roots-chinas-crimes-against-humanity-targeting-uyghurs-and-other-turkic-muslims/.

[77] Stephenie Nebehay, *China rejects geocide charge in Xinjiang, says door open to U.N.*, REUTERS (Feb. 22, 2021), https://www.reuters.com/article/us-china-rights/china-rejects-genocide-charge-in-xinjiang-says-door-open-to-u-n-idUSKBN2AM1UX.

[78] Albert Zhang, et al., *Strange bedfellows on Xinjiang: The CCP, fringe media, and US social media platforms,* AUSTL. STRATEGIC POLI, INST. (March 2021) https://s3-ap-southeast-2.amazonaws.com/ad-aspi/2021-03/Strange%20bedfellows.pdf?VersionId=mOh5mC5B_a08J6ntNwTC2q6GdjtWz4di.

[79] CGTN Documentary, This is Xinjiang Episode 1 Pathways, YOUTUBE (April 6, 2021), https://www.youtube.com/watch?v=RK13DpnKzCU.

[80] James Griffiths, *From cover-up to propaganda blitz: China's attempts toc control the narrative on Xinjiang*, CNN (April 27, 2021), https://www.cnn.com/2021/04/16/china/beijing-xinjiang-uyghurs-propaganda-intl-hnk-dst/index.html.

[81] *Id.*

region are supported by the Uyghur population.[82] Chinese diplomats and state media also have amplified false information from fringe media outlets. For example, Grayzone, a fringe news source which has accused the United States government of making up the Xinjiang forced labor stories to drive a wedge between the U.S. and China, has been cited over 300 times in State news outlets.[83] In its report on disinformation in Xinjiang, the Australian Strategic Policy Institute suggested that "the CCP will continue to trial new ways of distracting from, and suppressing where it can, international criticism of its systems of governance and control."[84]

IV. National Security Measures under International Law

Though the disinformation campaigns related to the Hong Kong National Security Law and the suppression of the Uyghur population bear many factual differences, there are some overarching similarities. In both of the situations, the Chinese government has deliberately used disinformation to justify the enactment of a controversial policy carried out under the auspices of national security. In Hong Kong, the law was enacted to ensure that those who threatened the security of Hong Kong would be punished, and in Xinjiang, the labor camps have been justified by a need to prevent extremism and terrorism, which are vital to China's national security interests. This Part will aim to answer the question of whether these policies are legal under international law and what the legality of these measures means for the future of disinformation in China.

International law confers sovereignty on each country, meaning that each state has a "presumptive duty… to respect the outcome of political processes internal to the others."[85] The Tallin Manual notes that "a state may not intervene, including by cyber means, in the internal or

[82] *See* Zhang, *supra* note 31.
[83] *Id*.
[84] *Id*.
[85] Brad R. Roth, *The Enduring Significance of State Sovereignty*, 56 FLA. L. REV. 1017 (2004).

external affairs of another state."[86] Recently, many countries have enacted national security legislation to combat information warfare which threatens sovereignty, political independence, and the territorial integrity of states.[87] This type of legislation protects against foreign interference and meddling in internal affairs, particularly through the dissemination of information online.

Under these new laws and international standards, it is not clear whether China is in violation of international law. China would likely argue that its information campaigns abroad are not intended to meddle in the internal affairs of other countries, but rather to "correct" the narrative surrounding its own internal affairs in the global discourse. Whether China is in violation of Hong Kong's sovereignty depends on the future of "one country, two systems." China certainly sees Hong Kong as part of China, although Hong Kong does have its own representation in some multilateral organizations like the WTO and APEC. It is therefore unclear whether attempts to cast China's disinformation activities in Hong Kong as illegal will be persuasive.

There is, however, another important international legal basis for evaluating these disinformation campaigns and national security measures. Many critics argue that China is in violation of international human rights law in both Hong Kong and Xinjiang. In Hong Kong, the Chinese government has been accused of using an excessively broad definition of "endangering national security" to suppress essential freedoms.[88] Specifically, the law entails prosecution of individuals for exercising their rights to free expression, association, and assembly, which is in violations of Hong Kong and Beijing's commitments under international human rights law in the

[86] Tomoko Nagasako, *Global disinformation campaigns and legal challenges*, 1 nt. CYBERSECUR. L. REV. 125 (2020).
[87] *See* Zhang, *supra* note 31.
[88] *Hong Kong: National Security Law has created a human rights emergency*, AMNESTY INTERNATIONAL (June 30, 2021), https://www.amnesty.org/en/latest/press-release/2021/06/hong-kong-national-security-law-has-created-a-human-rights-emergency/.

International Covenant on Civil and Political Rights.[89] While the ICCPR allows expression to be restricted for the protection of national security and public order, some might argue that China has gone too far in interpreting this provision. Furthermore, the situation in Xinjiang has prompted many observers to call it a crime against humanity, which, under the Rome Statute of the International Criminal Court, is defined as "offenses that are knowingly committed as part of a widespread or systematic attack against any civilian population."[90]

In response to these claims, China has called upon countries to respect its sovereignty and follow the principles of non-interference. Said China's Foreign Ministry spokesperson, Zhao Lijian, has said "[t]he affairs of Hong Kong, Xinjiang and Tibet are China's internal affairs, and the outside world should not interfere. All should abide by the purposes and principles of the UN Charter… oppose politically motivated and groundless accusations against China based on disinformation, and oppose interference in China's internal affairs under the pretext of human rights."[91] This statement not only speaks to China's view that these measures are legal, but it also reflects how it uses disinformation as a tool to deflect international criticism and create an alternate narrative to justify its controversial policies.

V. Conclusion

China's state disinformation apparatus touts the successes of the government which has stoked a sense of nationalism, inspiring millions of citizens to use their own platforms to fervently defend the Chinese government and endorse false narratives in the global discourse. China's

[89] Thomas Kellogg & Lydia Wong, *Hong Kong's National Security Law: A Human Rights and Rule of Law Analysis,* CENTER FOR ASIAN LAW OF GEORGETOWN LAW, February 23, 2021, https://www.law.georgetown.edu/law-asia/wp-content/uploads/sites/31/2021/02/GT-HK-Report-Accessible.pdf.
[90] *See* Schaack & Wang, *supra* note 31.
[91] *Foreign Ministry Spokesperson's Remarks on the Statement by Friendly Countries in Support of China at the 47th Session of the Human Rights Council*, CONSULATE GENERAL OF THE PEOPLE'S REPUBLIC OF CHINA IN NEW YORK (June 22, 2021), https://www.fmprc.gov.cn/ce/cgny/eng/mtsw/fyrth/t1885874.htm.

disinformation is therefore both a stabilizing force domestically and a destabilizing force abroad. This use of disinformation is concerning, given that it strategically uses it to justify its most controversial policies and restrict essential freedoms. Through analyzing the current status of international law to address disinformation it is clear that China will ardently defend itself by invoking the fact it has sovereignty over its internal affairs. There is therefore a need for the international community to further codify permissible uses of state disinformation and to more clearly define the line between national security and freedom of speech.

Disinformation and Information Warfare under International Law

I. Introduction

There is a modern-day conflict being waged, but instead of using bullets and bombs, modern countries are waging a war of words. Nations across the world have invested in and developed their online security and network technologies in disparate ways but still generally allow for information to flow freely across borders through the internet.[1] Simultaneously, the globe has experienced a rapid divergence in the factual character of information being shared online. In the wake of these developments, the global community has experienced widespread online misinformation and disinformation over the last two decades.

Using rhetoric as a figurative battering ram is a war tactic dating back centuries. Fortunately, today's international law condemns information warfare outside of sanctioned conflict.[2] Yet despite international laws, Russia and the U.S. have had a well-documented tension over "disinformation" and information influence operations since the advent of the internet.[3] With the dawn of social media and the newfound ease of online media dissemination, there is a new urgency to define the rules of engagement for foreign information and disinformation exchanges.

Current international conflict law does not fully account for information operations that influence domestic civilian populations through social media.[4] The international framework thus

[1] *Shaping Europe's Future: Open Internet,* European Commission (2021), https://digital-strategy.ec.europa.eu/en/policies/open-internet.
[2] Council of the EU, *EU imposes the first ever sanctions against cyber-attacks*, European Council: Press Releases (Jul. 30, 2020), https://www.consilium.europa.eu/en/press/press-releases/2020/07/30/eu-imposes-the-first-ever-sanctions-against-cyber-attacks/.
[3] Douglas Selvage, *Operation "Denver": KGB and Stasi Disinformation regarding AIDS*, Wilson Ctr. (Jul. 22, 2019), https://www.wilsoncenter.org/blog-post/operation-denver-kgb-and-stasi-disinformation-regarding-aids.
[4] Anna-Maria Talihärm, *Towards Cyberpeace: Managing Cyberwar Through International Cooperation,* UN Chronicle (2013), https://www.un.org/en/chronicle/article/towards-cyberpeace-managing-cyberwar-through-international-cooperation.

allows nations such as Russia to enjoy freedom of movement online to affect their international agenda. Without a narrower scope of internet laws and norms to account for modern "information warfare," nations that protect free and open speech online will be vulnerable to adverse impact from foreign disinformation tactics. This section specifically focuses on state-sponsored information operations in order to isolate the threshold for internationally prohibited information actions. Particular scrutiny will be applied to the common definitions of disinformation, misinformation, and information warfare in order to provide context to the current state of US and Russian information operations on the global internet.

II. A Vague Definition for Disinformation and Information Operations

Misinformation is generally defined as "false information, or [the] dissemination of such information not necessarily in the knowledge that it is false."[5] In contrast, disinformation is defined by Merriam-Webster as "false information deliberately and often covertly spread (as by the planting of rumors) in order to influence public opinion or obscure the truth."[6] However, the European Commission expanded the definition in its 2018 report to include "verifiably false or *misleading* information"[7] (emphasis added). Despite universal agreement that "intent to deceive" is a critical element separating disinformation from misinformation, there is disagreement on whether disinformation includes "misleading" communications.

Despite dissonance on what exactly qualifies as "disinformation," there is general acceptance on what information warfare entails. "Information warfare" is generally defined as "a strategy for the use and management of information to pursue a competitive advantage, including

[5] *Misinformation,* Merriam-Webster.com, https://www.merriam-webster.com/dictionary/misinformation (last visited 01 Dec, 2021).
[6] *Disinformation,* Merriam-Webster.com, https://www.merriam-webster.com/dictionary/disinformation (last visited 01 Dec, 2021).
[7] Commission Proposal for a European Approach to Tackling online disinformation, at 3, COM (Apr. 26, 2018).

both offensive and defensive operations."[8] Nearly every nation that has participated in an armed conflict within the last half century has also employed a form of disinformation against foreign adversaries through information warfare.

Information warfare was rampant during the Cold War among both Soviet and Western powers in the late 20th century. Regardless, the term "information warfare" is curiously absent from any modern official United States Government definitions.[9] The term is additionally not used to describe actions taken by the Russian government, although "information warfare" is often used by Russia to describe foreign state's adversarial actions in the information environment .[10] Instead, the U.S. military employs the term "Information Operations," which it defines as the "integrated employment of the core capabilities of electronic warfare, computer network operations, psychological operations, military deception, and operations security, in concert with specified supporting and related capabilities, to influence, disrupt, corrupt or usurp adversarial human and automated decision making while protecting [the U.S.'s] own."[11]

Similarly, Russia has abstained from using the term "information warfare," and taken to the term "information confrontation."[12] A 2021 NATO cybersecurity report claims Russia's Defense Ministry describes the information confrontation as "the clash of national interests and ideas, where superiority is sought by targeting the adversary's information infrastructure while protecting its own objects from similar influence."[13] Forms of disinformation take place in

[8] Catherine A. Theohary, CRS Report R45142, *Information Warfare: Issues for Congress* (Dec 01, 2021), https://crsreports.congress.gov/product/pdf/R/R45142/5.
[9] Catherine A. Theohary, *Defense Primer: Information Operations*, Cong. Rsch. Serv. (Dec 1, 2021), https://sgp.fas.org/crs/natsec/IF10771.pdf
[10] Janne Hakala, *RUSSIA'S STRATEGY IN CYBERSPACE*, NATO Strategic Communications Centre of Excellence (June 2021), https://stratcomcoe.org/cuploads/pfiles/Nato-Cyber-Report_15-06-2021.pdf
[11] JOINT CHIEFS OF STAFF, JOINT PUBLICATION 1-02: Department of Defense Dictionary of Military and Associated Terms (2016).
[12] Hakala, *supra* note 10.
[13] *See id.*

Russia's various information confrontation operations.[14] Russia is not restricted to participating in information confrontation abroad, and could just as easily employ their information operations within its own borders or against its own citizens[15].

Conversely, the United States has strict restrictions for the employment of information operations from both policy and legal standpoints. For example, the U.S. has claimed that it does not use "verifiably false" information in its Psychological Operations, which are a type of US information operation.[16] However, the U.S. Department of Defense defines Psychological Operations, or PSYOPS, as "planned operations to convey selected information and indicators to foreign audiences to influence their emotions, motives, objective reasoning, and ultimately the behavior of foreign governments, organizations, groups, and individuals… to induce or reinforce foreign attitudes and behavior favorable to the originator's objectives.[17]" While there may not be "false" information, the nature of PSYOPS employs misleading information, and thus classifies as disinformation according to the European Union. Unlike Russia, the United States only uses psychological operations abroad against foreign adversaries, and has strict legal codes which prevent U.S. PSYOP forces from targeting US citizens at any time, in any location globally, or under any circumstances.[18]

Despite their public condemnation of it, forms of disinformation are employed by both the US and Russia in their own versions of "information warfare." For the purposes of this section, "disinformation" will be understood to be the E.U.'s version of "verifiably false or misleading

[14] *See id.*
[15] Oreste Pollicino & Oleg Soldatov, *Striking the Balance Between Human Rights Online and State Security Concerns: The Russian Way in A Comparative Context*, 19 Ger. L.J. 85, 100 (2018).
[16] Marc D. Beaudreau & David Patrikarakos *War in 140 Characters: How Social Media Is Reshaping Conflict in the Twenty-First Century*, 72 Naval War College Rev. (2019).
[17] JOINT CHIEFS OF STAFF, JOINT PUBLICATION 1-02: Department of Defense Dictionary of Military and Associated Terms (2016).
[18] JOINT CHIEFS OF STAFF, JOINT PUBLICATION 3-13.2: Psychological Operations (2010).

information deliberately and often covertly spread in order to influence public opinion or obscure the truth." Additionally, the term "information operations" will be used to cover the full gamut of current modern information warfare, to include online "psychological operations," as well as "information confrontation." These narrowly scoped terms of disinformation and information operations will help address the current gaps between internationally prohibited and internationally permitted information conflicts.

III. Internationally Prohibited Disinformation Acts

There are three examples of information operations that both Russia and the United States understand to be illegal or prohibited by international norms. Three specific instances of internationally prohibited information operations are (1) the unsanctioned use of information operations as a means of force,[19] (2) the interference in a foreign nation's domestic affairs using information operations,[20] and (3) employing perfidy or treachery through information operations.[21] While there are other globally prohibited information crimes such as fraud, these three listed actions serve a critical function in limiting state-conducted adversarial information operations.

The UN Charter Article 2(4) provides that "All Members shall refrain in their international relations from the threat or use of force against the territorial integrity or political independence of any state, or in any other manner inconsistent with the Purposes of the United Nations.[22]" Accordingly, both the U.S. and Russia understand that online information operations which rise to the level of "force" are precluded by the UN Charter. Critically, there is no clear instance of when

[19] *See* U.N. Charter art. 2.
[20] *See Nicar. v. U.S.*, 1986 I.C.J. at 108, ¶ 205.
[21] *See* JOINT CHIEFS OF STAFF, JOINT PUBLICATION 3-13.2 at Legal Framework and Authorities: Psychological Operations (2010).
[22] U.N. Charter art. 2.

cyber information operations rise to the level of "force."[23] The Tallinn Manual, an academic, non-binding publication on the application of international law to cyber conflicts and warfare, addresses the execution of cyber operations. Rule 69 in the Tallinn Manual provides that in order for a cyber measure to qualify as a "use of force," the cyber operation would need to rise to the level of a "comparable non-cyber operation."[24] For example, if a state uses information operations to deceive city workers into shutting down an electrical grid using false and fraudulent messages or instructions, it would likely qualify as a prohibited use of force under the UN Charter.

The second form of prohibited disinformation would be any foreign engagements with another nation's civilian population that "interfere" with the sovereign jurisprudence of a nation.[25] In *Nicaragua v. United States*, the International Court of Justice defined a prohibited "intervention" as one that is "bearing on matters in which each State is permitted, by the principle of State sovereignty, to decide freely."[26] According to the Tallinn Manual, cyber "intervention" requires an element of coercion.[27] Although the International Court of Justice has stopped short of defining coercion, other courts in the United States have defined it as "the point at which pressure turns into compulsion."[28] Accordingly, any information operation which compels action by applying pressure through communications, such as threat or extortion, would be internationally prohibited. Cyber actions such as the deliberate interference in a nation's elections by affecting the vote would additionally classify as intervention. Yet crucially, influencing another nation's

[23] *See* Ashley C. Nicolas, *Taming the Trolls: The Need for an International Legal Framework to Regulate State Use of Disinformation on Social Media*, 107 Geo. L.J. Online 36, 43–44 (2018).
[24] TALLINN MANUAL 2.0 ON THE INTERNATIONAL LAW APPLICABLE TO CYBER OPERATIONS 491-96 (Michael N. Schmitt & Liis Vihul eds., 2017).
[25] *Nicar. v. U.S.*, 1986 I.C.J. at 108, ¶ 205.
[26] *Id.*
[27] TALLINN MANUAL 2.0 ON THE INTERNATIONAL LAW APPLICABLE TO CYBER OPERATIONS 491-96 (Michael N. Schmitt & Liis Vihul eds., 2017).
[28] Nicolas, *supra* note 23 at 43–44.

election using social media rhetoric, which does not contain the element of coercion, would be permitted under current international law.

Finally, the third form of disinformation which is internationally prohibited is treachery. Article 37 of the Geneva Convention prohibits perfidious messages, or using untrustworthy messages to deceive an adversary into a compromising position.[29] An example of this would be inviting a nation to engage in peaceful conversations, and then using treachery or perfidy to attack the unsuspecting nation during peace talks.

Prohibiting the use of disinformation as a means of force, foreign intervention, or treachery casts a wide umbrella. Nevertheless, there are still a number of other information operations which can still be legally conducted by states. The swath of permissible information operations enables today's current disinformation crises as discussed in the next section.

IV. Internationally Permitted Disinformation Acts

Outside of the prohibited international disinformation acts, there are two forms of "disinformation" that are actively sanctioned and used by the U.S. and Russia in their current operations. Namely they are (1) information operations during sanctioned armed conflicts,[30] and (2) psychological operations during peacetime.[31]

In the last forty years the United States has used information to affect the information environment surrounding its enemies. During the Gulf War, the United States deployed a leaflet campaign containing a slew of persuasive messages in order to persuade Iraqi troops to surrender.[32] Additionally, Army PSYOPS teams have used multimedia products such as leaflets, radio

[29] *See* Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I), art. 37, June 8, 1977, 1125 U.N.T.S. 3.
[30] *See* 10 U.S.C.A. § 394.
[31] *See* JOINT CHIEFS OF STAFF, JOINT PUBLICATION 3-13.2 at Legal Framework and Authorities: Psychological Operations (2010).
[32] *See* Beaudreau & Patrikarakos, *supra* note 16.

broadcasts, and social networking to shape Afghanistan populations into believing that both al-Qaeda and the Taliban do not live up to Islamic ideologies.[33] Although the U.S. claims it never uses verifiably false information in these information operations, they still qualify as disinformation under the EU definition.

Soviet Cold War-era disinformation campaigns attempted to convince the international community into supporting Soviet foreign policy. For example, their use of the 1985 Christian Peace Conference in Prague provided a messaging platform to convince global communities that the Soviet invasion of Afghanistan had been justified.[34] Both U.S. information operations in the Middle East and Soviet Cold War information operations classify as "disinformation" under the modern EU definition. Yet, due to the nature of their sanctioned armed conflicts at the time, the information operations were largely permitted by the international community.

The second form of permitted "disinformation" captures the crux of the current online disinformation crises. Namely, the modern use of public messaging to influence civilian populations is not precluded by international law. Rule 69 of the Tallinn Manual specifically explains that "non-destructive cyber psychological operations intended solely to undermine confidence in a government" would *not* qualify as a use of force precluded under the UN Charter.[35] While the Tallinn Manual is not binding law, it follows that the use of information to deceive and manipulate a civilian population on social media during peacetime falls short of a use of force and would be in accordance with both Article 2(4) of the UN Charter, Rule 69 of the manual, and the Geneva Convention.[36] As explained in the last section, the use of social media lacks the element

[33] *Id.*
[34] *See* Frank Goldstein Benjamin Findley, *Psychological Operations: Principles and Case Studies 160 (1996)*
[35] TALLINN MANUAL 2.0 ON THE INTERNATIONAL LAW APPLICABLE TO CYBER OPERATIONS 491-96 (Michael N. Schmitt & Liis Vihul eds., 2017).
[36] *See* Nicolas, *supra* note 23 at 49–50.

of coercion which would otherwise classify it as "intervention" precluded by the International Court of Justice. Correspondingly, US and Russian information operations conducted outside of active armed conflicts now benefit from glaring absence of terminology such as "warfare" or "force" in their official definitions.

Modern examples of peacetime influence operations include Russia's social media messaging in the Ukraine. For the last seven years, Russian-backed social media posts have disseminated pro-Russian claims across online Ukrainian social media spaces.[37] Examples include social media posts claiming Ukraine was a center of right-wing pro-Russian political extremism that was widely supported by the political parties of Ukraine.[38] The intent of these messages was to garner support for right-wing extremism that was already prevalent in national political discourse. In reality, Ukraine's nationalist parties are relatively weak compared to other EU nations, however, the slightly misleading disinformation was effective in convincing some populations to support a pro-Russian political stance.[39] There is currently no international law preventing similarly "misleading" social media actions, especially considering that internet messages using "misleading" instead of "verifiably false" information does not qualify as disinformation under most policies. Their tactics are more likely to be classified as "propaganda" instead of "disinformation by most states, which completely bypasses international conflict laws as well as new "disinformation" policies.[40]

---

[37] *See* Sam Sokol, *Russian Disinformation Distorted Reality in Ukraine. Americans Should Take Note*, Foreign Policy, https://foreignpolicy.com/2019/08/02/russian-disinformation-distorted-reality-in-ukraine-americans-should-take-note-putin-mueller-elections-antisemitism/.
[38] *See* Andreas Umland, *The dangers of echoing Russian disinformation on Ukraine*, Atlantic Council, https://www.atlanticcouncil.org/blogs/ukrainealert/the-dangers-of-echoing-russian-disinformation-on-ukraine/.
[39] *See id.*
[40] *See Propaganda, Merriam-Webster.com,* https://www.merriam-webster.com/dictionary/propaganda (last visited 01 Dec, 2021).

It is clear that the current definition of "disinformation" falls short of capturing the full gamut of state-sanctioned information operations in both Russia and the U.S. that do not rise to an internationally prohibited level. While information warfare is somewhat accounted for in international law, the information actions such as social media psychological operations fall short of traditional "force" or "intervention." Consequently, states enjoy a legally permissive environment for information operations across most of the world's social media sites.

V. The Modern War of Words

If the ability to conduct information operations abroad is an offensive sword, then the ability to counter foreign information operations within a nation's own virtual borders is a defensive shield. Currently, nations such as the U.S. have a limited reach with their sword, and next to no shield to defend. Any nation or society that protects a free and open forum for communication online lacks the ability to defend or effectively counter-attack under the current international framework.

The United States is particularly susceptible to information operations on social media. Privacy law and protected speech cases such as *McIntyre v. Ohio Elections Commission* intersect in America to protect anonymous online speech, especially if it is political in nature.[41] Accordingly, the U.S. online environment makes it difficult to differentiate between domestic and foreign disinformation. An online "bot" is able to simply re-post disinformation that already exists on the internet and the US federal law does not provide a clear remedy to the affected civilian populations. Russia however, has a number of laws, such as Federal Law Numbers 242-FZ 398-FZ, 7-FKZ, 374-FZ and 375-FZ which allow for incredibly effective internal defense against

[41] *See McIntyre v. Ohio Elections Comm'n*, 514 U.S. 334 (1995).

foreign information operations.[42] Their laws allow for the active monitoring, identification, and censorship of private blogging sites, public media, and general online activity.[43] Accordingly, the Russian government's "shield" is effective at culling instances of information the government believes to be misleading or false.

While the US is able to legally conduct counter-influence operations, it is incredibly difficult to win a public information war online without escalating the form of engagement to impermissible levels of intervention or force. The current international framework allows for psychological operations, which are a form of warfare, to take place on global social media sites with steady consistency.[44] There is no international provision to currently shore the defenses against the new information "sword" that nations such as Russia swing with incredible accuracy.

VI. Conclusion

According to the US Senate Intelligence Committee on Russian Active Measures Campaigns and Interference in the 2016 US Election, disinformation is defined as "the intentional spread of false information to deceive.[45]" Yet many of the disinformation tactics employed by online "bots" simply use misleading but verifiably true information with the intent of deceiving a population and thus fly under the radar of many global internet policies and regulations..

The international community faces a problem because it struggles to define adversarial "disinformation" for what it is: a form of psychological warfare. The US cannot use psychological operations against their own population because it is a military operation. Psychological and information operations have a clear nexus with armed conflict and the use of force, yet the global

[42] *See* Pollicino & Soldatov, supra note 15 at 100.
[43] *See id.*
[44] *See* Theohary, *supra* note 8.
[45] S. Rpt. 116-290, *Russian Active Measures Campaigns and Interference in the 2016 US Election*(2020).

community allows for them to continue on social media sites daily simply because they don't fit the definition of "force," "intervention," "treachery," or even simple "disinformation." Without international consensus on what the rules of engagement are, online information operations against civilian populations will continue undeterred.

By redefining how the world defines information warfare and psychological operations, agreements can be signed to preclude the current virtual conflict. In order to reclaim peace, the world must first redefine war.

Differing Strategies for Disinformation by Authoritarian States and States' Potential Legal Responses to Them

I. Introduction

Propaganda, disinformation, surveillance, and censorship have long been tools of repressive regimes to maintain power domestically and foment unrest among their political adversaries. As authoritarian states such as China and Russia have evolved and developed their long-standing policies of authoritarian rule within their own borders to include the digital information technologies of the last few decades, they have steadily increased use of the same playbooks used domestically against foreign states. The approach has combined Cold War-era propaganda missions and techniques with modern technological efficiencies and outlets to inundate their adversaries in a form of constant information conflict. The strategies of China and Russia share emphasis on deepening social divisions and doubt in democratic regimes and capitalist markets, but differ somewhat in their tactics for doing so. China's technological advances have led it to highlight the potential role of artificial intelligence, including "deep fake" technology. While Russia appears to be more active today in its use of social media for spreading disinformation, China's activity has grown exponentially in recent years and likely utilizes a greater number of social media applications at this point.[1] Both nations remain confident in the ability of state-run news media, a staple of disinformation in the twentieth century, to effectuate

[1] Scott W. Harold et al., *Chinese Disinformation Efforts on Social Media*, RAND Corporation, 3 (2021), https://www.rand.org/pubs/research_reports/RR4373z3.html. The regional scope of China's social media disinformation efforts involves the use of many applications which are available only in certain Asian countries targeted by China. For example, Chinese military and civilian personnel are believed to have used China's WeChat, South Korea's Kakao Talk, and Vietnam's Zalo, in addition to American applications such as Facebook, Instagram, and Twitter present in regional states.

their aims in adversarial countries.[2] Indeed, Chinese state-run media outlets look at Russia's model as one to emulate going forward, describing RT (formerly Russia Today) as the Russian Federation's 'external propaganda aircraft carrier.'[3]

Existing regulatory powers and new legislative actions in the U.K. and Taiwan illustrate potential legal responses to the old and new disinformation strategies of China and Russia. The U.K. has responded to Russian disinformation campaigns by invoking previously seldom-used "due impartiality" regulations to sanction Russian media entities leading the campaigns. In Taiwan, a recently-passed act criminalized the use of disinformation to influence elections at all levels, primarily in response to previous Chinese efforts to do so.

Part II of this chapter will argue that Russia and China have adapted different tactics to effectuate similar strategies in propagating disinformation to their perceived adversarial nations. The Russian model follows a long history of propaganda and disinformation techniques honed in the Soviet era, but intensified through a new strategic vision and technological capabilities aimed mostly at Western nations. The Chinese model borrows from some of Russia's successes, including the synchronization of military and civilian assets, but relies more on manpower than technology in some cases and promulgates more pro-China disinformation than Russia does pro-Russia. China targets many of the same nations as Russia, but its primary target remains Taiwan. Part III of this chapter will discuss two examples of legislative and regulatory authority to combat these approaches. The long-standing telecommunications regulations of Ofcom, the U.K.'s telecommunications regulatory agency, offer a potential solution to meet disinformation that relies

[2] Alina Polyakova and Chris Meserole, *Exporting digital authoritarianism: The Russian and Chinese models*, Brookings Institution (August 2019), https://www.brookings.edu/wp-content/uploads/2019/08/FP_20190827_digital_authoritarianism_polyakova_meserole.pdf.
[3] Elizabeth Chen, *China Learning from Russia's "Emerging Great Power" Global Media Tactics*, 21 China Brief 2, 4 (Apr. 12, 2021).

on state-run media like the Russian model. Meanwhile, Taiwan's recent legislative push to fight Chinese disinformation in its election process is promising for a country affected strongly by social media disinformation in the Chinese model.

II. Comparing Russian and Chinese Disinformation Strategies

A. Russia

Russia's historical use of disinformation to disrupt its adversaries' political and economic processes and promote its own world vision is well-documented. The Soviet Union's "active measures" during the Cold War era consisted of coordinated efforts by Soviet intelligence agencies, other Soviet governmental organizations, and non-state actors to influence public opinion in foreign countries.[4] An illustrative example was the harmonized effort of Russian media and pro-Soviet foreign press in the 1980s to persuade the world, especially developing countries, that the United States had created the AIDS virus to use as a biological weapon.[5] Apprehensive about global public concerns over their rumored development of biological weapons, Soviet intelligence agencies worked with various newspaper and radio outlets to drown out the questions by encouraging overwhelming numbers of false reports about the United States' development of AIDS. In recent years, Russian efforts to sway the U.S. presidential campaign in 2016, United Kingdom referendum vote to leave the European Union in 2016, and Ukrainian public opinion during the secession of Crimea in 2014 have borne similar trademarks combining both intelligence agency and non-governmental actors' involvement, as well as overwhelming media influence. But

[4] Christina Nemr and William Gangware, *Weapons of Mass Distraction: Foreign State-Sponsored Disinformation in the Digital Age*, Park Advisors (March 2019), https://www.park-advisors.com/disinforeport.
[5] *Id*.

the potency of such efforts has been magnified by the speed with which disinformation may be disseminated and the exponentially greater number of outlets, primarily online, in modern times.[6]

Russia's recognition of the speed and power of modern technology in conjunction with its disinformation efforts has led it to transform what was formerly a joint intelligence and media operation into one that encompasses military principles, as well. This idea, solidified somewhat in a February 2013 article written by Russia's Chief of the General Staff, General Valery Gerasimov, likely grew out of Russian frustration with what Moscow views as an ideological defeat in the Cold War followed by its military's inability to convincingly oppose NATO forces in a conventional manner. Coalescing what it sees as propaganda successes by its intelligence agencies and favorable media during the Cold War with military prowess in relevant fields such as hacking and psychological operations, Russia's military and political leadership, led by Gerasimov, has built a new strategic framework for opposing its adversaries through disinformation.[7] The playbook calls for chaos as its overarching principle, which Gerasimov believes will achieve permanent unrest and conflict within enemy states.[8] Within this strategy, almost anyone sympathetic to Russia's goals is a potential actor. Actors range from traditional actors such as media outlets and intelligence agencies to Russian citizens with hacking capabilities, psychological warfare soldiers, and massive networks of bots that can sow disorder with little to no human oversight.[9] Russian bot networks are generally given human direction, but post or retweet on social media and message boards automatically, often attempting to make Russian

[6] Christopher Paul and Miriam Matthews, *The Russian "Firehose of Falsehood" Propaganda Model*, RAND Corporation, 2 (2016), https://www.rand.org/pubs/perspectives/PE198.html.
[7] Molly McKew, *The Gerasimov Doctrine*, POLITICO Magazine (September/October 2017), https://www.politico.com/magazine/story/2017/09/05/gerasimov-doctrine-russia-foreign-policy-215538/.
[8] *Id*.
[9] Dmitry Volchek and Daisy Sindelar, *One Professional Russian Troll Tells All*, Radio Free Europe/Radio Liberty, March 25, 2015. "According to a former paid Russian Internet troll, the trolls are on duty 24 hours a day, in 12-hour shifts, and each has a daily quota of 135 posted comments of at least 200 characters."

disinformation "trend" when it would not otherwise do so.[10] Likewise, anyone or any entity can become a target of Russia's strategy within its adversary nations, as Russia's ultimate goal is not to bolster its own position, but rather to weaken that of its enemies. Moscow is not content to repeat the mistakes of its past, where attempts to oppose the West on military terms were ultimately negated by Western economic growth and political stability.[11] Its new doctrine aims to destabilize such strengths first by creating confusion and polarization within the electorate and population of its adversaries.

The command and control structure of Russia's modern disinformation is housed within the military leadership, as well as the Kremlin and other high-level political actors. General direction on where, when, and to what degree new disinformation campaigns are to take place moves from the highest levels of the Russian government into its first level of organizations and proxies, primarily encompassing attributed broadcast media like Sputnik and RT, as well as unattributed groups like the Internet Research Agency, the Russian "troll farm" responsible for sowing discord in the 2016 U.S. presidential campaign.[12] These organizations operate with varying levels of independence from Moscow, but altogether form the implementation arm of the disinformation chain, utilizing their social media platforms, fake and real accounts associated with such platforms, and traditional online, radio, and print media to amplify the message conveyed by Moscow.[13] Often, local media and private social media users further amplify the messaging by continuing to spread messaging originally featured on the proxies organizations' accounts. The events in this chain are not always linear, and often feature several nodes operating at the same

[10] Elizabeth Bodine-Baron et al., *Countering Russian Social Media Influence*, RAND Corporation, 10 (2018).
[11] Alina Polyakova and Chris Meserole, *Exporting digital authoritarianism: The Russian and Chinese models*, Brookings Institution (August 2019), https://www.brookings.edu/wp-content/uploads/2019/08/FP_20190827_digital_authoritarianism_polyakova_meserole.pdf.
[12] Bodine-Baron, *supra* note 10 at 7.
[13] *Id*. at 10.

time, but their direction is always aimed at infecting as many communications channels, user accounts, and messaging services as possible.[14] Russian strategy boils down to an attempt to proliferate disinformation rapidly and repetitively using its chain of actors, without great concern for plausibility in its claims, which causes chaos for its adversaries.

B. China

Unlike its neighbor to the north, China does not have an extensive history of using disinformation to disrupt adversary political and economic processes, but it has gained tremendous experience in this arena over the last several years. One significant difference between the two authoritarian nations remains China's desire to foster pro-China sentiments in target nations. Whereas Russia's focus, particularly through its use of traditional state-run media and social media campaigns, tends to be almost exclusively focused on fomenting discontent, China's policy splits efforts between creating discontent and attempting to foster pro-China sentiment in the target population. Following a military reorganization of the People's Liberation Army (PLA) in 2015, including the creation of a new military branch, the PLA Strategic Support Force (PLASSF), China has also signaled a new approach to information warfare, one that more clearly embraces social media disinformation campaigns.[15] This integration of military leaders and assets to be employed in spreading disinformation that originates from the highest political and civilian levels of the Chinese Communist Party and its media apparatus looks increasingly similar to Russia's infrastructure. One difference of note is that despite numerous studies on the effectiveness of Russian disinformation campaigns and a growing admiration for aspects of the Russian approach, China does not yet appear to employ bots as a major component of its social media efforts like

[14] Paul, *supra* note 6 at 2.
[15] Scott W. Harold et al., *Chinese Disinformation Efforts on Social Media*, RAND Corporation, 3 (2021), https://www.rand.org/pubs/research_reports/RR4373z3.html.

Russia.[16] Instead, there is evidence that China has utilized its population advantages to spread its propaganda, including over 20 million volunteers, many of whom are members of the Communist Youth League or are enrolled in Chinese universities.[17]

The highest intensity disinformation efforts by China in the last decade have focused on Taiwan, which has become a test-bed for Chinese disinformation strategies, and best illustrates its current model. The Chinese playbook in Taiwan seeks to create discord and division among rival political factions, diminish trust in the government, increase a sense of isolation in Taiwan and abandonment by its allies, and present China as an irresistible foe that offers a prosperous future.[18] For example, following the 2018 nine-in-one local elections in Taiwan, Taiwanese government officials from the National Security Bureau implicated PLASSF as the primary organization responsible for social media disinformation campaigns that disrupted the elections.[19] Experts in Taiwan believe China is propagating up to 2,400 separate pieces of disinformation on a daily basis, possibly higher during election seasons.[20] Engineers working under PLASSF published an article just before the elections describing the doctrine by which they could effectively manipulate Taiwanese society in general.[21] The strategy called for coordinated military and civil efforts, and described the enormous potential impact that "local agents," Chinese nationals or sympathizers fluent in Taiwanese dialect and culture, could have on social media disinformation spread.[22] Other accounts note that Chinese authorities paid multiple Taiwanese media groups to cover Chinese issues in a positive manner and stop mentioning issues such as the Tiananmen Square crackdown

[16] Harold, *supra* note 15 at 28.
[17] Ryan Fedasiuk, *A Different Kind of Army: The Militarization of China's Internet Trolls*, The Jamestown Foundation, 21 China Brief 8 (Apr. 12, 2021).
[18] *Id*. at 4.
[19] Nathan Beauchamp-Mustafaga and Jessica Drun, *Exploring Chinese Military Thinking on Social Media Manipulation Against Taiwan*, The Jamestown Foundation, 21 China Brief 25, 26 (Apr. 12, 2021).
[20] Harold, *supra* note 15 at 65.
[21] Harold, *supra* note 15 at 57.
[22] *Id*.

of 1989, believed to assist Taiwanese political parties that are "softer" on China issues.[23] Though never proven, some Taiwanese media claim the disinformation campaign was strong enough to swing the election away from the ruling Democratic Progressive Party (DPP) that year toward the more pro-unification Chinese Nationalist Party (KMT). China's increasing combination of joint military-civil disinformation strategy draws it closer to that of Russia, but it has yet to focus the bulk of its attacks on Western countries as Russia does. Also, while both countries' strategies involve voluminous disinformation attacks to create chaos, China attempts to do so in some cases by painting itself in a positive light.

III. The Potential and Limits of Regulation and Legislation for State Disinformation

Policy recommendations for countering disinformation campaigns by Russia and China are far-ranging and broad in scope, with legal or regulatory actions encompassing just one potential method, alongside population education efforts, self-censoring social media companies, and public fact-checking initiatives. Experimental research in psychology shows that in all likelihood, a multi-pronged approach will be necessary to combat the highly effective models employed by Russia and China.[24] The efficacy of legal or regulatory action is likely to be greatest in preventing state disinformation from entering a target state's media and social media networks at the outset, or dissuading such attempts through the potential for criminal prosecution. Once such disinformation has made its way into a target population, education and fact-checking initiatives could help stem the flow of the "infected" information. Several laws and regulations passed by various target states

[23] Yimou Lee and I-hwa Cheng, *Paid 'news': China using Taiwan media to win hearts and minds on island*, Reuters, August 9, 2019,https://www.reuters.com/article/us-taiwan-china-media-insight/paid-news-china-using-taiwan-media-to-win-hearts-and-minds-on-island-sources-idUSKCN1UZ0I4.
[24] Christopher Paul and Miriam Matthews, *The Russian "Firehose of Falsehood" Propaganda Model*, RAND Corporation, 9 (2016), https://www.rand.org/pubs/perspectives/PE198.html.

of Russian and Chinese disinformation in recent years can illuminate the potential impact that legal and regulatory action might have against state disinformation.

A. Telecommunications Regulation

One recent case in the United Kingdom demonstrates the potential for use of existing telecommunications regulatory schemes to hold traditional state-run media companies accountable for the spread of disinformation.[25] The U.K.'s Office of Communications ("Ofcom") prevailed in England's High Court against RT on a claim the latter had breached its longstanding "due impartiality" provisions of the Communications Act of 2003 (2003 Act) in news and television programs.[26] Ofcom's original sanction against RT called for a fine of £200,000 (approximately $271,000) for violating section 320 of the 2003 Act on seven occasions between March and May 2018.[27] According to Rule 5.11 of section 320 of the 2003 Act, "due impartiality must be preserved on matters of major political and industrial controversy. . . by the person providing a service . . . in each programme."[28] The High Court upheld the sanction, finding RT had violated "due impartiality" by carrying stories on the poisoning of a former Russian intelligence officer in England, U.S. activities in Syria, and Nazi sympathizers in Ukraine that did not "accurately and adequately" reflect alternative views or opinions on the matters.[29] The High Court further noted the requirement by all licensees to comply with "due impartiality" as a strict condition of their licenses, which could theoretically be revoked if patterns of disinformation continued.

[25] *R (Autonomous Non-Profit Organisation TV-Novosti) v. The Office of Communications*, [2020] EWHC (QB) 689 (Eng.).
[26] *Id*. at 2.
[27] *Broadcast Standards Cases*, 369 Ofcom Broadcast and On Demand Bulletin 6 (Dec. 20, 2018), https://www.ofcom.org.uk/__data/assets/pdf_file/0020/131159/Issue-369-Broadcast-and-On-Demand-Bulletin.pdf.
[28] *Id*.
[29] *Id*. at 23.

The effectiveness of the 2003 Act and sanctions against RT in this case are not without limits. While there do not appear to be further instances of sanctions brought by Ofcom since the decision, indicating RT may have curbed its use of disinformation on broadcast media, the 2003 Act is limited in its scope. As the High Court noted, RT was not prevented from broadcasting any material, even that with highly-biased views, so long as there was balance to ensure "due partiality" and alternative views.[30] Further, under British law, the 2003 Act and other regulatory statutes do not prohibit RT from publishing "lawful" political or news programs online or via social media without needing to satisfy the "due impartiality" requirement.[31] The "due impartiality" requirement demands that news be reported with due accuracy, and mistakes acknowledged and corrected on air quickly, with political allegiances by person presenting such news declared.[32] Such a narrow regulation affecting only broadcast media might be more effective in states whose populations do not see social or online media as a "substitute for the immediacy and impact" of broadcast media.[33] In a country such as Taiwan, in which a very high percentage of the population is believed to consume news via social media, the effect of a similar act would likely be muted.[34]

Telecommunications regulators in nations where broadcast media does form a majority of popular news consumption could put in place similar rules or regulations to the 2003 Act, though political practicalities could make this a difficult legislative endeavor. For example, the U.S.'s Federal Communications Commission (FCC) has issued several public statements in recent years that it does not have the statutory authority to revoke broadcaster licenses based on the content of

[30] *Id*. at 22.
[31] *Id*.
[32] *See* Rules 5.1-5.3, Section Five: Due Impartiality and Due Accuracy, *Ofcom Broadcasting Code* (Jan. 05, 2021), https://www.ofcom.org.uk/tv-radio-and-on-demand/broadcast-codes/broadcast-code/section-five-due-impartiality-accuracy.
[33] *Id*.
[34] Harold, *supra* note 1 at 53 (citing several studies on Taiwanese social media penetration that show Facebook, as one example, reaches between 89 and 97 percent of Taiwanese internet users).

broadcasted material.[35] In fact, the FCC does possess the authority to issue fines under the broadcast hoax rule and revoke broadcast companies' licenses under the news distortion policy, but only in incredibly rare cases has the FCC even considered their application and even more infrequently found violations of them.[36]

Ofcom's decision to leverage its sanction power and not revoke RT's license entirely was at the regulatory discretion of the telecommunications agency, and the High Court found it to be within Ofcom's statutory authority and necessarily "proportional" to the "actual or potential" harm caused.[37] Given its warnings for three previous violations of the "due impartiality" requirement from 2012-2016 without repercussions, Ofcom deemed the substantial sanction to be adequate, but did not rule out a future revocation of RT's broadcast license if it continued to breach the requirements. One of the greatest limitations of current telecommunications regulators in the fight against state disinformation is their lack of rules or authorities for executing the same penalties against social media disinformation violators that they may against broadcast violators.

B. Targeted Anti-Disinformation Legislation

Facing the prospect of a potentially overwhelming Chinese disinformation campaign on the eve of its presidential election in January 2020, the Taiwanese parliament passed the Anti-Infiltration Act two weeks before the scheduled election. In addition to public education efforts on disinformation, cyber security defense initiatives, and policies encouraging self-censoring of fake accounts by social media companies such as Twitter and Facebook, the Anti-Infiltration Act was

[35] Jon Brodkin, *Ajit Pai refuses Democrats' request to revoke Sinclair broadcast licenses*, ars Technica (Apr. 12, 2018), https://arstechnica.com/tech-policy/2018/04/ajit-pai-refuses-democrats-request-to-revoke-sinclair-broadcast-licenses/.
[36] Joel Timmer, *Broadcasters and Trump's False Information on Coronavirus: What Role for the FCC?*, Just Security (Apr. 27, 2020), https://www.justsecurity.org/69843/broadcasters-and-trumps-false-information-on-coronavirus-what-role-for-the-fcc/.
[37] Harold, *supra* note 1 at 30.

Taiwan's most direct recent action to combat China's disinformation efforts in its political process. The law criminalizes individuals convicted of spreading disinformation in various media, including social media, providing for up to five years in prison and authorizes fines up to roughly $330,600 for organizations or actors responsible for the spread of disinformation originating from outside Taiwan.[38] The legislation does not regulate the distribution of information since authorities may only prescribe sanctions after various government agencies are authorized to investigate and prove that foreign influence through disinformation has taken place. The purpose of the Anti-Infiltration Act was to prohibit and punish illegal foreign donations, canvassing, and lobbying to political parties, and the spreading of disinformation through any medium at the instruction of or with support from external forces hostile to Taiwan's democracy.[39] Taiwanese proponents of the law argue that previous regulations did not provide for strong enough mechanisms to oppose the disinformation strategies it routinely sees in recent years, including China's social media campaigns, and were only minimally effective against traditional forms of broadcast media engaged in disinformation. Critics of the law maintain that it detracts from Taiwan's historically strong free speech and free press traditions.

The Anti-Infiltration Law's initial effects were highly visible and lauded by the ruling Democratic Progressive Party and President Tsai Ing-wen following her successful re-election in the 2020 campaign. Immediately after enactment of the law, Master Chain, a pro-China media outlet accused of spreading Chinese propaganda and disinformation throughout Taiwan,

[38] Daniel Halpert, *Disinformation Prevention and Defending Democracy in Taiwan*, Brown Political Review (Apr. 8, 2020), https://brownpoliticalreview.org/2020/04/disinformation-prevention-and-defending-democracy-in-taiwan/.
[39] *Legislative Yuan Passes Anti-Infiltration Bill to Strengthen Defense for Democracy and Preserve Stable and Orderly Cross-Strait Exchanges*, Mainland Affairs Council, Republic of China (Taiwan), Press Release No. 101 (Dec. 31, 2019), https://www.mac.gov.tw/en/News_Content.aspx?n=2BA0753CBE348412&s=88E5E1EF1343B1B8#:%7E:text=The%20Anti%2DInfiltration%20Act%20simply,exchanges%20with%20peace%20of%20mind.

announced it was closing its operations in Taiwan and relocating to Beijing.[40] As an originally Taiwanese company, Master Chain was indicative of the sort of local, Taiwan-based intermediaries that Chinese PLASSF operations frequently target as distributors of disinformation in Taiwan. The Anti-Infiltration Act does provide a legal apparatus for punishing both the foreign state originator of disinformation and any local sympathizers that participate and may offer an effective counter-strategy for other target states facing similar circumstances. However, in its attempt to avoid disproportionately limiting freedom of expression and impose sanctions only after formal investigations by various executive agencies, the threshold for bringing punitive action is high. It requires a potentially drawn-out process with highly effective means for identifying and attributing the nefarious activity to a foreign state. This level of sophistication in the investigation aspect might prove too difficult for many states, particularly given the prowess of Chinese or Russian agents spreading disinformation through social media and leaving few clues of their involvement.

IV. Conclusion

Both Russian and Chinese disinformation strategies have shown themselves to be effective in impacting other countries' elections, among other activities. The two nations' strategies mimic one another in some respects, but diverge in others. Target countries of these attacks will need a comprehensive approach to combat them, but the legal responses by the U.K. and Taiwan offer some hope, if they can be tailored to respond to the specific strategies used by Russia and China. Western nations might see some success in utilizing their telecommunications regulators' powers against Russian state-sponsored media companies, while Taiwan is better served constructing new

[40] Linda Zhang, *How to Counter China's Disinformation Campaign in Taiwan*, 100 Mil. Rev. 21, 25 (Sep.-Oct. 2020).

legislative powers that address the relatively newer disinformation threat on social media. In either case, all target nations will likely need both approaches and many more to meet the broad array of disinformation tactics employed by Russia and China now and in the future.

Social Movements and Democracy

In recent years, the spread of disinformation has led to democratic backsliding in countries around the world. This trajectory encompasses a host of concerns, including the lack of informed voting in elections, the polarization of factions in society, and the repression of free speech and expression rights. The democratization of the internet has led to a revolution in the ways that social movements engage with technology. More than ever before, the democratic process is shaped online.

While social movements have often emerged as a counter to regimes that seek to propagate disinformation, this top-down framework is not a monolith for understanding the role of disinformation in democratic societies. Social movements themselves have propagated disinformation in order to win elections or shape historical narratives. This Section, which examines case studies from Asia, North Africa, and Europe, reflects this inherent tension that complicates a coherent legal response.

Though there is no consensus on how to solve this complex issue, these case studies highlight the greatest dangers and point to areas for reform including strengthening international legal protections, regulating social media, and promoting digital literacy.

Disinformation Campaigns in the COVID-19 Era: Combatting Russian COVID-19 Disinformation in the European Union

I. Introduction

In January 2020, when the World Health Organization (WHO) declared an outbreak of a novel Coronavirus in Wuhan, China a Public Health Emergency of International Concern, it is hard to imagine anyone foresaw the extent to which the COVID-19 virus would impact the world. As of this paper, nearly five million people have been killed by COVID-19, and the WHO has confirmed more than two hundred million cases.[1] Globally, the response to the pandemic has involved more than sixteen trillion dollars in fiscal support.[2] Unsurprisingly then, responses to the pandemic, especially government handling of vaccine programs, have become a critical issue in both domestic and international politics.[3]

The COVID-19 pandemic has also seen disinformation and misinformation take a central position on the global stage. Antivax protests have erupted around the world catalyzed by inaccurate claims including those that vaccines contain tracking chips, cause recipients to become magnetized, or cause impotence and testicular swelling.[4] These and other examples of misinformation have dominated social media platforms and news outlets. In the United States,

[1] *See* World Health Organisation, *WHO Coronavirus (COVID-19) Dashboard*, World Health Organization (Oct. 27, 2021), https://covid19.who.int/
[2] *See* Federico Filippini & Eduardo Levy Yeyati, *Social and Economic Impact of COVID-19,* Brookings (June 8, 2021), https://www.brookings.edu/research/social-and-economic-impact-of-covid-19/
[3] In March 2021, Slovakian Prime Minister Igor Matovic resigned over a deal to purchase two million Russian Sputnik V vaccines despite no approval by the European Medical Agency, which is required for all vaccines in the EU. In September 2021, California State governor Gavin Newsome faced a highly publicized recall election due to his handling of the pandemic, including mask and vaccine mandates.
[4] *See* Nicki Minaj (@NICKIMINAJ), Twitter, September 13, 2021, https://twitter.com/NICKIMINAJ/status/1437532566945341441

dozens of people have been hospitalized for taking the horse deworming drug Ivermectin to cure their COVID-19 symptoms due to publicized and uninformed claims of its effectiveness.[5]

State actors have taken advantage of the pandemic to implement targeted disinformation campaigns aimed at fomenting dissent in foreign states, improving the image of their handling of the pandemic, and shifting blame for the outbreak of the virus. Russia, a familiar and well-versed actor in weaponizing and distributing disinformation, has developed extensive COVID-19 disinformation campaigns directed at the West, that criticize their handling of the pandemic, spread disinformation about Western-developed vaccines, and promote its own Sputnik V vaccine. The two largest targets of these campaigns are the United States and the European Union (EU). In the EU the goal of spreading COVID-19 disinformation is to sow distrust of the EU government, and create and widen fissures between member states, particularly between the Eastern and Western blocs. The Eastern member states who have had historically closer relationships with Russia and its president, Vladimir Putin, are also a prime potential market for the Russian Sputnik V vaccine. While the EU has developed policies and programs aimed at combatting disinformation, specifically with an eye towards Russia in the past few years, the COVID-19 pandemic has highlighted the limitations of these legal frameworks.

This paper aims to analyze the strengths and limitations of EU law in combatting Russian state-lead disinformation campaigns surrounding the COVID-19 pandemic. Part II provides a brief background on Russian COVID-19 disinformation campaigns in Europe. Part III lays out existing legal frameworks in the EU at supernational and national levels and analyzes some weaknesses in

[5] *See* Franny White, *Five Oregonian's Hospitalized Due to Misuse of Ivermectin for COVID-19,* OHSU (Sep. 17, 2021), *https*://news.ohsu.edu/2021/09/17/five-oregonians-hospitalized-due-to-misuse-of-ivermectin-for-covid-19#:~:text=The%20Oregon%20Poison%20Center%20has,to%20an%20intensive%20care%20unit

these approaches. Part IV suggests potential ways to improve existing EU law as well as alternative approaches based on national approaches and international case law.

II. Russian COVID-19 Disinformation Strategy

Russia has long been among the most prolific and successful state actors in spreading disinformation targeting foreign states using social media. Russian disinformation notably played a significant role in the 2016 United States Presidential Election, the 2016 Dutch Referendum, the French *Gilet Jaune* protests, and Brexit.[6] Given the global nature of the pandemic, publicized failures in containing the virus, and contentious and politicized debates surrounding vaccines, COVID-19 has provided an ideal arena for Russian disinformation campaigns to sow dissent in the EU among member states' governments over their handling of the pandemic and vaccine rollouts.

Russian COVID-19 disinformation falls broadly into three categories: 1) shifting responsibility for the virus onto Western governments; 2) critiquing Western government responses to the virus while extolling the successes of Russian COVID-19 measures; and 3) casting doubt onto Western vaccines and spreading claims as to the efficacy of its own Sputnik V vaccine. In targeting the EU and its members' responses to the pandemic, these campaigns aim to discredit the EU government and thus improve its own image by comparison, domestically and in the global theater. Furthermore, by delegitimizing Western vaccines, such as Pfizer's, as well as the European Medical Agency (EMA), which is responsible for approving vaccines in the EU, Russia aims to sow distrust towards EU governments and foster support for its own Sputnik V vaccine program.

[6] *See* Agnieszka Legucka, *Russia's Long-Term Campaign of Disinformation in Europe,* Carnegie Europe (March 19, 2020), https://carnegieeurope.eu/strategiceurope/81322

As of October 2021, the EU's External Action Service (EEAS) had identified almost one thousand Russian media articles spreading pro-Kremlin disinformation related to the COVID-19 pandemic.[7] Examples of Russian disinformation that the EU has flagged include media stories stating that the Sputnik V vaccine is the most effective vaccine and that the EU is preventing its use in Europe for political reasons,[8] that the EU is intentionally spreading disinformation about the Sputnik V vaccine as an attack to the Russian government,[9] and that Western vaccines are designed to kill recipients to reduce global populations.[10] Disinformation is particularly concentrated around discrediting the EMA for not approving the Sputnik V vaccine for political reasons to attack Russia,[11] which as a result, has led to vaccine hesitancy in Russia and difficulties achieving widespread vaccination rates in parts of the EU.

EU leaders have critiqued Russian state-controlled media for intentionally spreading disinformation with the goal to promote use of the Russian Sputnik V vaccine.[12] In April 2021, the EEAS issued a report detailing Russian COVID-19 disinformation campaigns that aimed to foment discontent in the EU and promote its own Sputnik V vaccine.[13] The report found that "foreign state actors have sensationalized and misrepresented information about the safety of Western-made

[7] *See* EUvsDisinfo, *Disinfo Database,* EUvsDisinfo (Oct. 27, 2021), https://euvsdisinfo.eu/disinformation-cases/?date=&per_page=
[8] *See* Yelena Karajeva, *Vaccine War: How Europe Fights Sputnik V,* Sputnik News (Aug. 7, 2021), https://sputniknews.lt/20210707/vakcinu-karas-kaip-europa-kovoja-su-sputnik-v-17399375.html
[9] *See* Sputnik News, *The Director of Russian Intelligence Reveals European Smear Campaigns Against His Country's Vaccines,* Sputnik News (May 18, 2021), https://sptnkne.ws/Gq3N
[10] *See* Shaban Syed, *US Instigated Hybrid War Against China and Russia Evolves Into a "Deadly" Vaccine War,* Geopolitica.ru (July 7, 2021), https://www.geopolitica.ru/en/article/us-instigated-hybrid-war-against-china-and-russia-evolves-deadly-vaccine-war
[11] *See* EUvsDisinfo, *EEAS Special Report Update: Short Assessment of Narratives and Disinformation Around the COVID-19 Pandemic* at 1 (April 28, 2021), https://euvsdisinfo.eu/eeas-special-report-update-short-assessment-of-narratives-and-disinformation-around-the-covid-19-pandemic-update-december-2020-april-2021/
[12] *See* Josep Borell, *The Essential Fight Against Disinformation and Manipulation,* European Union External Action Service (Dec. 27, 2020), https://eeas.europa.eu/headquarters/headquarters-homepage/91038/essential-fight-against-disinformation-and-manipulation
[13] *See* EUvsDisinfo, *EEAS Special Report Update*

vaccines and fueled anti-vaccination movement within the EU."[14] While measuring the direct impact of disinformation campaigns is almost impossible, mistrust in the EU and its reaction to the pandemic has undoubtedly grown over the past year. In Italy, the European country hit hardest by the pandemic, a 2020 poll found that 88% of respondents believed that the EU was not doing anything to help Italy during the crisis, and at the same time, support for Russia surged significantly between 2019 and 2020.[15] Hungary independently granted emergency authorization to the vaccine, while Slovakian prime minister Igor Matovic secretly imported two hundred thousand doses of the Sputnik V with the intent to purchase two million doses, despite the EMA's pending approval and lack of agreement within his government.[16] Recognizing the danger of social media and fake news campaigns politically and medically, at the beginning of the pandemic the EU centered combatting disinformation as one of the key pillars of its COVID-19 strategy.

III. Legal Frameworks

The current framework for combatting disinformation in the EU is balanced between the guarantees of freedom of expression enshrined in the founding documents of the EU and the need to prevent the spread of disinformation. On one side, the Charter of Fundamental Rights of the European Union guarantees everyone "the right to freedom of expression… to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers."[17] Similarly, Article 10 of the European Convention on Human Rights (ECHR) guarantees everyone the right to freedom of expression.[18] The European Court of Human Rights

[14] *Id.* at 3
[15] *See* Julia Baer-Bader, *EU Response to Disinformation from Russia on COVID-19,* DGAP (June 2, 2020), https://dgap.org/en/research/publications/eu-response-disinformation-russia-covid-19
[16] *See* Kevin Connolly, *Sputnik V: How Russia's Covid Vaccine is Dividing Europe,* BBC (April 17, 2021), https://www.bbc.com/news/world-europe-56735931
[17] Charter of Fundamental Rights of the European Union, art. 11
[18] *See* European Convention on Human Rights, art. 10

(ECtHR) has interpreted this to be a positive obligation on countries to create a favorable environment for participation in public debate.[19]

There are, however, widely recognized exemptions from freedom of speech protections in both European and International law. In the ECHR, Restrictions under Article 10 are permissible if they comply with three conditions: They must be (1) prescribed by law, (2) introduced for protection of one of the listed legitimate aims, and (3) necessary in a democratic society. Legitimate grounds that could justify interference include national security, territorial integrity or public safety, and the prevention of disorder or crime.[20] Similarly under International Law, the International covenant on Civil and Political Rights ("ICCPR") Article 19(3) recognizes that restrictions may be imposed on any form of expression or means of its dissemination 'as are provided by law and are necessary: (a) for respect of the rights or reputations of others; (b) for the protection of national security or of public order (*ordre public*), or of public health or morals."[21] The restrictions must also be the least intrusive instrument among those which might achieve the desired result.

On the other side of the balance is the EU's desire to combat disinformation, in this case, specifically stemming from Russia. The prevalence of Russian disinformation campaigns over the past five years has caused the EU to state that "disinformation by the Russian Federation poses the greatest threat to the EU."[22] In 2015, the European Council created the East StratCom Task Force under the European External Action Service's (EEAS) Strategic Communications Division, specifically to deal with the problem of Russian disinformation in the EU. The East StratCom Task

[19] *See Dink v. Turkey*, 7124:09 Eur. Ct. H. R. (2010) (finding that Turkey violated the ECHR for failing to protect a journalist who was an outspoken member of the Armenian minority)
[20] ECHR art. 10 ¶2
[21] International Covenant on Civil and Political Rights, art. 19
[22] EU High Representative of the union for Foreign Affairs and Security Policy, *Action Plan Against Disinformation,* European Commission, JOIN(2018) 36 Final at 4 (Dec. 5, 2018)

Force functions primarily through its flagship project, EUvsDisinfo, which monitors and compiles examples of pro-Kremlin disinformation affecting the EU member states and surrounding countries. The aim of the program is "to increase public awareness and understanding of the Kremlin's disinformation operations, and to help citizens in Europe and beyond develop resistance to digital information and media manipulation."[23] Although EUvsDisinfo explicitly clarifies that it is not alleging that the examples it compiles indicate that the message is sponsored by the Russian state, many of the examples are from Russian state-controlled media organizations such as Sputnik and RT. In targeting disinformation, EUvsDisinfo has a two-step test: 1) the information must be verifiably false or misleading, according to the publicly available factual evidence and 2) it must originate in a Kremlin-funded media outlet or other information source that has clear links to the Russian Federation.[24]

The EU has also recognized the importance of self-regulation and co-regulation by and with social media companies, whose platforms have facilitated the spread of disinformation. In 2018, the EU created its Code of Practice on Disinformation (CPD), a nonbinding agreement voluntarily signed by major companies in the tech industry agreeing to self-regulate and combat disinformation on their platforms in five major areas. Facebook, Twitter, Mozilla, Microsoft, and recently TikTok, are all signatories of the CPD. The CPD sets out as one of its objectives, "ensur[ing] the integrity of services with regard to accounts whose purpose and intent is to spread disinformation."[25] Under the CPD, the signatories commit to developing policies and strategies for combatting misleading political advertising, enabling public disclosure of political advertising and ensuring that political advertisement are clearly distinguishable from editorial content, putting in

[23] EUvsDISINFO, *About*, EUvsDISINFO (Oct. 27, 2021), https://euvsdisinfo.eu/about/
[24] *See Id.*
[25] EU Code of Practice on Disinformation art. I, Apr. 2018

place measures to prevent misuse of automated bots, empowering consumers to identify and resist disinformation by investing in programs and technology, and supporting programs to monitor and research disinformation on their sites, including by independent fact checkers established by the European Commission.[26] Furthermore, under the best practices, signatories submit yearly self-reports setting out the state of play of the measures taken to comply with their commitments under the code.[27]

The European Commission also established its Action Plan Against Disinformation in 2018, which lays out a proposed foundation for the EU's policy to combat disinformation based on 1) improving EU capabilities to detect and expose disinformation; 2) strengthening coordination and responses to disinformation; 3) mobilizing the private sector to tackle disinformation; and 4) raising awareness and improving societal resilience.[28] The Plan establishes that member states "should support the creation of teams of multi-disciplinary fact-checkers and researchers… to detect and expose disinformation campaigns across different social networks and digital media."[29]

a. Weaknesses of Existing EU Policy

Despite these measures, Russian disinformation has continued to spread rapidly in the EU with few concrete actions taken to remove or eliminate the disinformation itself. The EUvsDisinfo program has faced significant critique in how it determines what constitutes disinformation, as well as the accuracy with which it is able to do so. In 2019 the site withdrew three Dutch articles it had identified as being Russian disinformation after backlash from the Netherlands and

[26] *See* EU Code of Practice on Disinformation arts. II(A)-II(E)
[27] *See Id.* at annex II
[28] *See Action Plan Against Disinformation* at 5
[29] *Id.* at 11

international journalistic organizations.[30] The articles it had identified as fake news were mis translated, or written as satire and identified as disinformation and uploaded to the EUvsDisinfo site.[31] The East StratCom Task Force is also constrained by a limited budget which limits its ability to sort through the immense number of articles that it needs to. Furthermore, due to EEAS policy, EUvsDisinfo cannot include any EU websites in its database, meaning that any disinformation originating from inside the EU is exempt from review.

The CPD also faces similar issues in enforcement and harmonization. There are distinct discrepancies in how signatory companies defined issues of disinformation or what constitutes "political" content, leading to no "consistent implementation of specific restrictions" of it.[32] The EU has recognized important policy decisions taken on by signatories in order to take action against malicious actors' use of manipulative techniques to boost the dissemination of disinformation and engage in collaborative activities with fact checkers and the research community.[33] However, it has also acknowledged that the Code suffers from hazy definitions and procedures, and lack of appropriate monitoring.[34] Furthermore, because the CPD is voluntary, there is no requirement that signatories take any action pursuant to the CPD or to their own internal policies.

---

[30] *See* Khan M. Peel, *EU Attack on Pro-Kremlin "Fake News" Takes a Hit,* Financial Times (May 6, 2021), https://www.ft.com/content/5ec2a204-3406-11e8-ae84-494103e73f7f

[31] *See Id.*

[32] Ethan Shattock, *Self-Regulation 2:0? A Critical Reflection of the European Fight Against Disinfromation*, Harvard Kennedy School (May 31, 2021), https://misinforeview.hks.harvard.edu/article/self-regulation-20-a-critical-reflection-of-the-european-fight-against-disinformation/

[33] *See* European Commission, *Assessment of the Code of Practice on Disinformation- Achievements and Areas for Further Improvement,* European Commission (Sep. 19, 2020), https://digital-strategy.ec.europa.eu/en/library/assessment-code-practice-disinformation-achievements-and-areas-further-improvement

[34] *See Id.*

b. National Approaches

Given the lack of a centralized EU approach to disinformation, member states have developed a variety of approaches independently. Many EU member nations have domestic laws prohibiting the spread of disinformation, which range from laws specifically criminalizing election manipulation to laws permitting the government to completely block access to content arbitrarily considered fake news.[35] Using two examples from the EU, it is clear that the existing frameworks at a supernational level in the EU have led to inconsistent and conflicting approaches that are incapable of combatting disinformation across the entirety of the EU.

In 2018, in response to concerns over disinformation targeting elections, France passed Law 2018-1202 of December 22, 2018 on the Fight Against the Manipulation of Information providing that courts could order an online platform to remove misleading or untrue statements that aimed to influence the election in the three months beforehand.[36] The law applies only to platforms with more than five million monthly users, and requires that any challenged speech be 1) manifestly false, 2) widely spread in an artificial manner, and 3) aimed at disrupting public peace or the integrity of a ballot. The law further allows the *Conseil Supérieur de l'audiovisuel*, the French broadcasting agency, to suspend television channels controlled by or under the influence of a foreign state if they deliberately spread disinformation to impact the election.[37]

[35] In 2020, Hungary passed a law aimed at combatting COVID-19 disinformation that included provisions criminalizing spreading misinformation regarding COVID-19 with no clear standards by which to define what misinformation is. For further discussion see Ronan O Fathaigh & Joris Van Hoboken, *Regulating Disinformation in Europe: Implications for Speech and Privacy* at 19, UC Irvine Journal of International, Transnational, and Comparative Law VI art. 3 (May 2021)

[36] *See* Loi 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information [Law 2018-12-2 of December 22, 2018 on the Fight Against the Manipulation of Information], Journal Officiel de la République Française [Official Gazette of France], No. 0297

[37] *See Id.* at art. 6

In 2018, Germany passed the Network Enforcement Act (NetzDG) to combat hate speech and other illegal forms of speech on social media platforms.[38] The law requires large social media platforms with more than two million registered users to promptly remove "manifestly illegal" content as defined by twenty-two provisions of the criminal code or face significant criminal fines.[39] Criminalized speech includes incitement to racial and religious violence and child pornography, as well as dissemination of propaganda material of unconstitutional organizations and defamation of the state and its symbols.[40] The law does not provide for any judicial oversight, and has been used to censor political art, satire, and statements of political party leaders,[41] leading to criticism from Human Rights Watch, the Global Network Institute, and the UN Special Rapporteur on Freedom of Opinion and Expression[42] among others.

IV. Potential Solutions

In December 2020, the European Commission also proposed the Digital Services Act (DSA), establishing rules governing EU digital spaces and designating a Digital Services Coordinator in each member state. Among the requirements of the DSA is that tech companies promptly remove "illegal content" that has been reported on their sites by the relevant national judicial or administrative authorities, or risk significant fines.[43] The DSA would also require social media companies to implement measures to allow users to flag disinformation and cooperate with

[38] *Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken*
[39] *See* Network Enforcement Act § 3, Fed. Law. Gaz. I, 33522 (2017)
[40] *See German Criminal Code* § 86
[41] *See* Human Rights Watch, *Germany: Flawed Social Media Law,* Human Rights Watch (Feb. 14, 2018), https://www.hrw.org/news/2018/02/14/germany-flawed-social-media-law
[42] *See* Mandate of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, OL DEU 1/2017, (June 1 2017) (expressing that the restrictions under the NetzDG bill are incompatible with Article 19 of the ICCPR because they are too vague); Global Network Initiative, *Proposed German Legislation Threatens Free Expression Around the World,* Global Network Initiative (April 20, 2017), https://globalnetworkinitiative.org/proposed-german-legislation-threatens-free-expression-around-the-world/
[43] *See* Proposal for a Regulation on a Single Market for Digital Services (Digital Services Act) art. 5(b), COM/2020/825 (Dec. 15, 2020)

"trusted flaggers" to identify and take down such content.[44] A significant issue with the DSA, however, is that "illegal content" is dependent on member states' law,[45] and with regard to disinformation, there are drastically different concepts of disinformation which makes cooperation and harmonization of combatting disinformation incredibly difficult. Nonetheless, the DSA provides a potential framework for taking affirmative action to remove disinformation, particularly regarding coordinated disinformation campaigns propagated by state actors like Russia. In the context of COVID-19, the DSA is particularly important because it specifically targets social media companies, which is where COVID-19 disinformation finds footholds among groups who already distrust modern medicine and particularly vaccinations.[46] As discussed above, existing frameworks such as EUvsDisinfo are limited in that they only address news outlets, and in particular ones that are less reputable and more likely to publish Russian disinformation, without addressing situations in which these news articles are subsequently widely shared on social media. Therefore, the DSA offers a way to fight disinformation where it ends up influencing people- on social media- rather than where it originates, in a Russian troll farm for example.

While there are myriad implications for freedom of speech in any disinformation regulating approach, there is already some precedent in European courts for monitoring illegal online content and taking actions to combat it. In *Glawischnig-Piesczek v. Facebook*, the Court of Justice of the European Union established that injunction against content hosts could be enforced globally when illegal content was identified on their sites, and that "identical" of "equivalent" content could be

[44] *See* European Commission, *The Digital Services Act: Ensuring a Safe and Accountable Online Environment,* , European Commission (Oct. 27, 2021), https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en#new-obligations
[45] *See* Digital Services Act, art. 2
[46] For a further discussion of anti-vaccination disinformation and social media see Nicola Biller-Andorno & Federico Germani, *The Anti-Vaccination Infodemic on Social Media: A Behavioral Analysis*, PLoS One (March 2021)

ordered to be similarly removed.[47] The court elaborated that "equivalent" content would require that the "information contain specific elements which are properly identified in the injunction, such as the name of the person concerned… the circumstances in which that infringement was determined and equivalent content to that which was declared illegal."[48] This standard carefully refrains from putting any onus on technology companies to conduct independent assessments of hosted information, but rather requires that the injunction identify with a high level of specificity of the illegal content that the host must remove.

It is impossible for any EU program, regardless of budget or number of "trusted flaggers," to sort through the entirety of Russian disinformation on the internet. Nonetheless the EUvsDisinfo is a promising program in that it exclusively targets disinformation linked to the Russian state and is further limited to news sources rather than posts on social media platforms. Drawing from *Glawischnig-Piesczek*, the EU could criminalize certain acts of disinformation, at least those stemming from a foreign government aimed at intentionally destabilizing the EU and use analysts under either EUvsDisinfo or the DSA to identify those limited instances of disinformation, and remove them from the entirety of applicable social media platforms. This approach would avoid the issue of fragmented domestic approaches, as a centralized identification of Russian disinformation in Brussels issued by a court, would be cause to remove the illegal disinformation across all member states.

Given the increasing pervasiveness of disinformation, particularly in the digital sphere, it is clearly a legitimate interest of the EU to protect itself and its member states from targeted campaigns that have drastic impacts on public safety. In the case of Russian disinformation, where the intent is to cause separations between EU states, restrictions on certain speech can also be

[47] *See Eva Glawischnig-Piesczek v. Facebook Ireland Limited*, C-18/18 Eur. Ct. H.R. at 14 (2019)
[48] *Id.* at 45

justified as being for the protection of territorial integrity. As such, the above framework is consistent with exceptions to freedom of speech protection under the ICCPR and the ECHR.

V. Conclusion

Because the EU is composed of twenty-seven states representing a wide range of political positions, a top-down EU approach is already contentious and fragmenting, and any action would face immediate challenges under freedom of speech protections. Nonetheless, leaving disinformation legislation to individual states has led to fractured approaches from member states, and ineffective half measures at the EU level encouraging self-regulation and lacking enforceability. The DSA is an important step forward in regulating the increasingly prevalent issue of disinformation, particularly in instances such as the COVID-19 Pandemic, where foreign states use targeted campaigns across media platforms. These regulations, however, must be implemented within a larger framework that harmonizes what disinformation is and empowers EU organs and courts to actively remove state disinformation that threatens the EU's public safety and territorial integrity.

Popular Movements and Disinformation: A Case Study on Hong Kong

I. Introduction

In June of 2019, Hong Kong residents took to the streets to protest a proposed Extradition Bill, with an estimated one to two million participants at public events.[49] Demonstrations continued throughout the year and into 2020, and the situation of the movement today remains unclear.[50] The protests were met with force, as reports of police brutality and arrests of demonstrators inundating the media.[51] While demonstrators attracted international attention from foreign governments, the United Nations, and NGOs, their efforts ultimately resulted in passage of the National Security Law. The consequences of the movement are still playing out, and the world continues to watch as Beijing grips tighter control over Hong Kong.

During the protests, a viral tweet from an unverified user showing video footage of a tank at a train station warned of Chinese military interference at the Hong Kong-Shenzhen border.[52] However, the train station sign says "Longyan," a city in Fujian hundreds of miles away.[53] Similarly, a photograph of a woman who appears to be pregnant lying on the floor of a subway station following a mob attack was shared on social media with allegations that she had suffered a miscarriage after being attacked by other citizens following attendance at a protest.[54] This led to

---

[49] *See* Lukasz Zamecki, *"The Revolution of Our Times": Reasons for the Hong Kong Protests of 2019*, 6 CONTEMP. CHINESE POL. ECON. AND STRATEGIC RELS.: AN INT'L J., 899, 902 (2020).
[50] *See generally*, Chong Yiu Kwong, *Hong Kong, a Truly International City in 2019/2020: Timeline of Incidents – International and Human Rights Perspectives*, 6 CONTEMP. CHINESE POL. ECON. AND STRATEGIC RELS.: AN INT'L J., 833 (2020).
[51] *See generally*, *Hong Kong Protests: Updates and Latest on City's Political Unrest*, CNN, https://www.cnn.com/specials/asia/hong-kong-protests-intl-hnk, (last visited Oct. 1, 2021) (providing up to date information on the Hong Kong protests).
[52] @c338ki_selina, TWITTER (June 5, 2019, 9:38 AM), https://twitter.com/c338ki_selina/status/1136265963857297408; Jessie Yeung, *Hong Kong Isn't Just Battling on the Streets: There is Also a War on Misinformation Online*, CNN (Aug. 11, 2019 7:47 PM), https://www.cnn.com/2019/08/11/asia/hong-kong-fake-news-intl-hnk/index.html.
[53] *See* Yeung, *supra* note 4.
[54] *See* Yeung, *supra* note 4.

public outrage against the police for failure to protect protestors from violence.[55] Others claimed the woman was not pregnant at all, and the truth remains unknown.[56] These two examples demonstrate the ways in which disinformation permeated among the people of Hong Kong during the movement.

Part I discusses the role of the internet and social media in protests in Hong Kong, both in the 2014 Umbrella Movement and the 2019-2020 events, which allowed disinformation to spread easily. Part II shows the domestic law response to both disinformation and the protests. Part III it highlights the international legal response and constraints to the domestic laws.

The media has often addressed how the Chinese government spread disinformation about the protests. For example, *The New York Times* reported that China has been using state and social media to brand "demonstrations as a prelude to terrorism" by manipulating the context of images and videos.[57] Facebook, Twitter, and Google even announced plans to take action against the disinformation about the protests.[58] These issues are indeed serious; the role of the state and disinformation is discussed in other Chapters of this paper. Without downplaying the issues with state-sponsored disinformation, this Chapter seeks to address the role disinformation played among protestors and activists and the legal frameworks available to address this disinformation.

II. The Internet, Social Media, and Protests

This Part addresses how the political situation in Hong Kong created an environment ripe for disinformation to spread. Any evidence regarding the spread is largely anecdotal, and it is

[55] *See* Yeung, *supra* note 4.
[56] *See* Yeung, *supra* note 4.
[57] Steven Lee Myers & Paul Mozur, *China is Waging a Disinformation War Against Hong Kong Protesters*, N.Y. TIMES (Aug. 13, 2019), https://www.nytimes.com/2019/08/13/world/asia/hong-kong-protests-china.html.
[58] Emily Stewart, *How China used Facebook, Twitter, and YouTube to Spread Disinformation about the Hong Kong Protests*, VOX (Aug. 23, 2019), https://www.vox.com/recode/2019/8/20/20813660/china-facebook-twitter-hong-kong-protests-social-media.

beyond the scope of this Part to attempt to quantify it. Instead, this Part draws from the characteristics of the popular movements in the region to show how and why social media played a role in the protests, enabling disinformation to spread. It discusses the Umbrella Movement, which was the first major popular movement in Hong Kong where widespread use of the internet was utilized to galvanize the movement. The Umbrella Movement is relevant to the most recent protests because it shows how use of the internet contributed to the movement. Additionally, the structure of the 2019–2020 protests evolved from the Umbrella Movement.

Hong Kong has seen a number of protests over the past two decades, evolving from organized rallies with agendas, press briefings, and mass mobilization to spontaneous and decentralized organization.[59] Recent protests can be characterized by occupation, a process that "disrupts the established order by disrupting traffic, work patterns, and the very idea of ordinariness."[60] This is what happened in 2014, when protesters occupied areas around the Government Headquarters, highways, and pedestrian pathways calling for democratic elections.[61] The Umbrella Movement was an "occupation protest for 'genuine democracy' conducted by Hong Kong people from late September to mid-December 2014."[62] Unlike previous protests, it was the first time Chinese society "boast[ed] an advanced use of [i]nternet technology engaging in large-scale public protest."[63] Therefore, the Umbrella Movement provides a useful backdrop into understanding the role of the internet and social media in the subsequent 2019-2020 protests, which created the opportunities for disinformation to play a role.

---

[59] *See* Edmund W. Cheng, *Street Politics in a Hybrid Regime: The Diffusion of Political Activism in a Post-Colonial Hong* Kong, 226 THE CHINA QUARTERLY 382, 394 (2016).
[60] *Id.*
[61] FRANCIS L. F. LEE & JOSEPH M. CHAN, MEDIA AND PROTEST LOGICS IN THE DIGITAL ERA: THE UMBRELLA MOVEMENT IN HONG KONG, 3 (2018) (ebook).
[62] *Id.* at 1.
[63] Bryan Druzin & Jessica Li, *The Art of Nailing Jell-o to the Wall: Reassessing the Political Power of the Internet*, 24 J.L. & POL'Y 1, 5 (2015).

The Umbrella Movement began through formal organization by the so-called Occupy Central Trio—two professors and a reverend—who had recruited volunteers, organized deliberation days, and engaged in discourse on civil disobedience.[64] Six months later, what was intended to be a disciplined occupation resulted in tens of thousands of citizens occupying the district where the Government Headquarters are located in September 2014, sparked by the government's decision to maintain the pro-China nomination committee responsible for endorsing candidates for the upcoming 2017 election.[65] These events met forceful resistance from police, and images of tear gas fired at protestors protecting themselves with umbrellas quickly captivated the world.[66] The Umbrella Movement then "evolved into a range of improvisational, decentralized, and even individualized actions both within and outside the occupied areas, and both online and offline."[67]

These actions included use of social media and alternative media sites to amplify voices and spread coverage of contentious episodes throughout the movement. For example, on social media individuals could change profile pictures to yellow umbrellas in support of the movement or blue ribbons in support of police.[68] Additionally, ordinary people engaged in "citizen journalism, which emphasizes street-level perceptions, native accounts and alternative voices," to cover the events.[69] These alternative media sites "attracted new writers and readers and further blurred the boundaries between objective observers and passionate participants."[70] Social media and the internet, then, played an important role in the Umbrella Movement. And in the 2019–2020 protests, it continued to play a role, since now the people of Hong Kong had experience using it to

[64] *See* LEE & CHAN, *supra* note 13, at 3.
[65] *See* LEE & CHAN, *supra* note 13, at 4.
[66] *See* LEE & CHAN, *supra* note 13, at 4.
[67] *See* LEE & CHAN, *supra* note 13, at 5.
[68] *See* LEE & CHAN, *supra* note 13, at 5.
[69] Cheng, *supra* note 11, at 397.
[70] Cheng, *supra* note 11, at 398.

support and expend popular movements. And with use of social media comes opportunities for disinformation to thrive.

In the summer of 2019, the people of Hong Kong again took to the streets. They protested proposed legislation that would allow extradition of Hong Kong residents to the People's Republic of China.[71] More broadly, the protests demonstrated the fears that the "one country, two systems" principle was degrading into one unified country governed by Beijing.[72] The 2019–2020 protests were characterized by disorganization and decentralization combined with the widespread use of social media and the internet, and a growing radicalization and acceptance for violence. These characteristics made the situation conducive to the spread of disinformation.

Unlike the Umbrella Movement, these protests were entirely decentralized.[73] Protestors did not seek to occupy a single place so as to avoid being dispersed by law enforcement, and they similarly lacked a single organization or leader to whom they could point.[74] As such, social media played a role in "linking up the movement's decentrali[z]ed structure and decision-making processes."[75] Protest participants passed information through *LIHKG* (the Hong Kong based forum website similar to Reddit), Airdrop disks, and Bluetooth.[76] Additionally, protestors achieved worldwide visibility by maintaining contact with politicians abroad, collecting money, and promoting slogans online.[77]

The decentralization of the protests also promoted radicalization of the demonstrations.[78] Crowds included participants wearing masks and goggles to be prepared for physical confrontation

[71] *See* Zamecki, *supra* note 1, at 900.
[72] *See* Zamecki, *supra* note 1, at 900.
[73] *See* Zamecki, *supra* note 1, at 930.
[74] *See* Zamecki, *supra* note 1, at 904–05.
[75] Zamecki, *supra* note 1, at 931.
[76] Zamecki, *supra* note 1, at 931.
[77] Zamecki, *supra* note 1, at 933.
[78] Zamecki, *supra* note 1, at 905.

with the police and paramedics on hand for when, or if, violence broke out.[79] While a number of complex reasons beyond the scope of this Chapter caused this radicalization, one large catalyst was the police's violent reaction to the protestors on June 12.[80] In response to a small group of young protestors throwing bricks, water bottles, and umbrellas at police among a group of a largely peaceful, unarmed protestors, police fired tear gas and rubber bullets into the crowd.[81] Videos showed protestors being beaten and stampedes running to escape.[82] A number of other violent incidents also went viral and shook the city and the world.

The tense situation where real violence was occurring between both sides shows both why disinformation playing off those truths could spread and was believable. It also shows why its spread may not have changed the outcome of the movement; real things were happening alongside the disinformation, so it is unclear how much the disinformation perpetuated existing tensions. As discussed in the introduction, reports of disinformation spreading on social media have surfaced. For example, allegations spread online that a 22-year-old student was chased, or even pushed, fatally off the edge of a parking garage by police, and police then blocked the ambulance from reaching the student.[83] These allegations were apparently unsubstantiated, yet led to protestors clashing with police over the death.[84] It is unclear whether the disinformation ultimately changed the outcome of the movement (or whether the protests produced an outcome at all), but it did empower the government to take further action in opposition to the goals of the movement, as discussed in the next Part.

---

[79] Zamecki, *supra* note 1, at 933.
[80] Zamecki, *supra* note 1, at 933.
[81] *See* Preeti Jha, *Hong Kong Protests: The Flashpoints in a Year of Anger*, BBC NEWS (Aug. 31, 2020) https://www.bbc.com/news/world-asia-china-53942295.
[82] *See id.*
[83] Shelly Banjo & Natalie Lung, *Fake News, Rumour Stoking Division in Hong Kong*, NAT'L POST; DON MILLS, ONT., Nov. 15, 2019, at 9, ProQuest document ID 2314688082.
[84] *See id.*

III. The Domestic Law Response: The Police Doxxing Case, the National Security Law, and a Potential Fake News Law

One response to disinformation occurring throughout the protest movement came from the courts. In November 2019, the Court of First Instance[85] addressed "doxxing" of police officers and their families.[86] Doxxing, a form of cyber-bullying, involves sharing personal data of individuals online without their consent that is intended to intimidate or harass them.[87] It also includes assisting or instigating others to commit these acts.[88] In the police doxxing case, a group on the platform Telegram with nearly 200,000 members leaked personal information of police and their families members, including contact information, residential addresses, Hong Kong ID card numbers, car registration plate numbers, and personal particulars of family members.[89] The public nuisance action was brought by the Secretary for Justice to enjoin any individuals from conducting these actions. The court noted the effect of the injunction order on restricting certain fundamental rights, but determined that, when weighing the effect against the "rights of police officers and their family members to respect and privacy, as well as the need to maintain public order," there was "clear utility in the reminder of the risks to the maintenance and the application of the rule of law in Hong Kong."[90]

Notably, the court rejected a "media exemption" proposed by the Hong Kong Journalists Association ("HKJA"). HKJA feared the injunction would impact journalistic activity by

[85] The Court of First Instance is part of the High Court and has appellate and original jurisdiction over criminal and civil matters. It hears appeals from criminal cases heard in Magistrates' Courts and from tribunals cases. It has original jurisdiction over most serious criminal offenses has unlimited jurisdiction over all civil matters. *See Hong Kong Judiciary – High Court*, JUDICIARY (Dec. 2018), https://www.judiciary.hk/en/court_services_facilities/hc.html.
[86] Sec'y for Just. v. Pers. Unlawfully and Willfully Conducting Etc., [2019] 5 HKLRD 500 (C.F.I.).
[87] *Id.* at 503.
[88] *Id.*
[89] *Id.* at 504–05.
[90] *Id.* at 501.

restricting lawful day-to-day duties of journalists and deterring informants from disclosing information to the press that might cause a police officer to be pestered or interfered with by investigative journalism.[91] The court agreed that adding the exemption could be beneficial in differentiating between "lawful journalistic activity and fake journalists whose activity" is not "news activity," as "fake news activity is not news activity."[92] However, the court determined that the exemption would not be "necessary or helpful" because "[s]imply disclosing personal data to a data user" is not prohibited under the injunction.[93]

Beyond its discussion of the media exemption, the court did not explicitly address disinformation. However, the decision is notable because it acknowledged the problem with fake news and determined that the maintenance of public order outweighed protecting freedom of speech and expression. *The China Daily*, a state-sponsored news organization, reported on the decision as "necessary to get a grip on the protests."[94] It stated that the decision "reveals the lack of legal tools to contain the violence stemming from mistrust and misdirection of the public by fake news with a view to defaming police and destabilizing society."[95] The article called for the Hong Kong government to enact a law against fake news.[96]

Less than a year later, on June 30, 2020, the National People's Congress passed the Law of the People's Republic of China on Safeguarding National Security in the Hong Kong Special Administrative Region ("National Security Law").[97] The National Security Law was enacted for

---

[91] *Id.* at 511.
[92] *Id.* at 513.
[93] *Id.* at 514.
[94] Raymond Li, *HK Should Consider Law Against Fake News*, CHINA DAILY GLOB. (Nov. 15, 2019, 7:25 am), chinadaily.com.cn/a/201911/15/WS5dcde255a310cf3e35577712.html.
[95] *Id.*
[96] *Id.*
[97] *Implementation Rules for Article 43 of the Law of the People's Republic of China on Safeguarding National Security in the Hong Kong Special Administrative Region Gazetted*, THE GOV'T OF THE HONG KONG SPECIAL ADMIN. REGION PRESS RELEASES (July 6, 2020, 21:51), https://www.info.gov.hk/gia/general/202007/06/P2020070600784.htm.

the purpose of "preventing, suppressing and imposing punishment for the offences of secession, subversion, organi[z]ation and perpetration of terrorist activities, and collusion with a foreign country or with external elements to endanger national security."[98] Hong Kong Chief Executive Carrie Lam then issued Implementation Rules to "improve enforcement mechanisms" pursuant to her powers under Article 43.[99] If police suspect that an "an electronic message published on an electronic platform" is likely to endanger national security, they are authorized to request the message publisher or platform to remove the message and/or restrict access to the message.[100] If the platform fails to remove the information, police can apply for a warrant to seize the electronic device and "take any action for removing that information as soon as practicable."[101]

The law does not directly address disinformation, but it does give the state tools to quell the spread of information—true or false—that "endangers national security." Over one hundred people have been arrested for violating the law, with more than four-fifths accused of political activities.[102] Those activities have included "displaying banners, posting in support of the city's independence on social media or organizing primary elections for legislative seats."[103] This broad measure may address disinformation, but the law has been heavily criticized by the international community, as discussed in the next Part. Nonetheless, Hong Kong lawmakers are currently considering implementing a "fake news" law to further regulate disinformation.[104] The legislation

---

[98] Zhonghuarenmingongheguo Xianggang Tebie Xingfang Qu Weihu Guojia Anquanfa (中华人民共和国香港特别行政区维护国家安全法) [Law of the People's Republic of China on Safeguarding National Security in the Hong Kong Special Administrative Region] (promulgated by Standing Committee of the National People's Congress, June 30, 2020, effective June 30, 2020), CLI.1.343624(EN) (Lawinfochina), at Ch. 1 Art. 1.
[99] *Supra* note 54.
[100] *Id.*
[101] *Id.*
[102] *See* Natalie Lung, Kari Soo Lindberg & Chloe Lo, *Hong Kong Makes 100th Arrest Using National Security Law*, BLOOMBERG, March 2, 2021, https://www.bloomberg.com/news/articles/2021-03-03/hong-kong-makes-100th-arrest-using-china-drafted-security-law.
[103] *Id.*
[104] Rhoda Kwan, *Hong Kong Gov't and Lawmakers Back 'Fake News" Law Plan; Press Union Chief Warns of New "Sword Over Journalists' Heads'*, H.K. FREE PRESS, July 21, 2021, https://hongkongfp.com/2021/07/21/hong-kong-govt-and-lawmakers-back-fake-news-law-plan-press-union-chief-warns-of-new-sword-over-journalists-heads/.

would "combat online information that did not amount to national security offen[s]es but was still capable of 'radicali[z]ing' the public."[105] If passed, the law would likely provoke a harsh response form the international community, as discussed in the next Part.

IV. The ICCPR and the International Criticism

The International Covenant on Civil and Political Rights guarantees the right to freedom of expression, which includes "freedom to seek, receive and impart information and ideas of all kinds . . . through any other media of his choice."[106] Article 4 of the National Security Law states that the "rights and freedoms, including the freedoms of speech, of the press, of publication, of association, of assembly, of procession and of demonstration" shall be protected, and cites to this treaty.[107] Nonetheless, the law has been heavily criticized by the international community and human rights groups. The UN human rights office has expressed that arrests of individuals under the National Security Law demonstrate that the law's overly broad offenses facilitate "abusive or arbitrary implementation," and it called on "authorities to uphold their obligations" under international law to refrain from using the law to "suppress the rights to freedom of expression, peaceful assembly and association."[108] Similarly, Amnesty International asserts that the law "has created a human rights emergency.[109] According to its research briefing, the law has enabled police to use "social media posts and online commentary as evidence that a person has 'endangered'

[105] *Id.*
[106] International Covenant on Civil and Political Rights art. 19(2), *opened for signature, ratification, and accession* Dec. 16, 1966, 999 U.N.T.S. 171 (entered into force Mar. 23, 1976).
[107] *Supra* note 50, at Ch. 1 Art. 4.
[108] *Hong Kong: UN Human Rights Office Urges Immediate Release of Arrested Activists*, UN NEWS (Jan. 7, 2021), https://news.un.org/en/story/2021/01/1081552.
[109] *Hong Kong: National Security Law Has Created a Human Rights Emergency*, AMNESTY INT'L (June 30, 2021, 5:01 am), https://www.amnesty.org/en/latest/press-release/2021/06/hong-kong-national-security-law-has-created-a-human-rights-emergency/.

national security."[110] A fake news law would likely meet the same response, as journalists have already objected to it as "one more weapon or sword" over their heads.[111]

Whether any of these measures have successfully combated disinformation by the popular movement is unknown. In the police doxxing case, the court observed its initial injunction did result in a "material and meaningful drop in the number of doxxing posts,"[112] but beyond that the answer is unclear. Similarly, the National Security Law may have been able to dispel the spread of disinformation, or at least caused the arrest of a number of people who may or may not have been responsible for the spread. Nonetheless, this all came with the cost of restricting the right to expression recognized by international law.

V. Conclusion

This Chapter has shown how disinformation could take root in the popular movements of Hong Kong, how Hong Kong domestic law attempted to counter it, and the constraints from international law.

[110] AMNESTY INT'L, HONG KONG: IN THE NAME OF NATIONAL SECURITY: HUMAN RIGHTS VIOLATIONS RELATED TO THE IMPLEMENTATION OF THE HONG KONG NATIONAL SECURITY LAW 8 (2021), https://www.amnesty.org/en/wp-content/uploads/2021/07/ASA1741972021ENGLISH.pdf.
[111] Kwan, *supra* note 61.
[112] Sec'y for Just. v. Pers. Unlawfully and Willfully Conducting Etc., [2019] 5 HKLRD 500, 510 (C.F.I.).

Disinformation and Democratic Transition in Tunisia

I. Introduction

It seems strange to recall how recently the internet and social media were heralded as the greatest hope for democracy in the 21st century. These were the tools that toppled tyrants, that amplified the voices of people long oppressed and connected grassroots movements around the world. That narrative has changed. In recent years, platforms like Facebook, Twitter, and YouTube have provided new hosts for an old pathogen: disinformation. As false information spread over social media has infected recent election cycles, the tendency of these platforms to promote disinformation has been identified as among the greatest threats to democracy today.

Perhaps the strongest example of the democratizing power of the internet began in the country of Tunisia. Its revolution in 2011 relied heavily on Facebook and other social media to generate a massive protest movement that toppled their dictator of 23 years. That movement then spread over social media to other countries in the region, setting off the wave of democratic protest movements known as the Arab Spring. But as Tunisia has navigated the rocky straits of a democratic transition, its earlier successes with social media have earned it no special respite from the spread of disinformation across those platforms.

This paper will examine Tunisia's struggles with disinformation in the context of its ongoing democratic transition. First, it will address the interaction between disinformation and freedom of expression. Next, will be a discussion of the development of freedom of expression in Tunisia, followed by a report on Tunisia's recent struggles with disinformation. Finally, the paper will evaluate the government's options for curbing disinformation, concluding that a civil society-focused approach provides a better solution for Tunisia's situation than a state-centric approach.

II. Disinformation and Freedom of Expression

Disinformation has proven especially difficult to control because it preys on a fundamental human right and a central element of democracy—freedom of expression. Under international human rights law (IHRL), freedom of expression entails not only the right to impart information and ideas to others, but also the right to seek and receive information.[1] Individuals may exercise these rights on any form of media, including the internet and social media platforms where most modern disinformation tends to spread.[2] IHRL permits restrictions on freedom of expression only in certain circumstances: to protect the rights or reputations of others, to protect national security, or to protect public order, health, or morals.[3]

Beyond its inherent value as a human right, freedom of expression also plays an indispensable role in any democratic system.[4] Established democracies tend to have strong norms, built over decades and centuries, that restrain governments from interfering with freedom of expression. By contrast, transitional democracies usually lack such deep-seated norms, and often struggle to strike the balance between freedom of expression and other matters of public concern, such as national security, public morals, or disinformation.[5] Building respect for freedom of expression is critical for the long-term success of a democratic transition.

Government action aimed at curbing disinformation must walk a tightrope to avoid undue infringement on freedom of expression. Governments have compelling and legitimate reasons to seek to control disinformation. Disinformation poses its own threats to human rights and

[1] International Covenant on Civil and Political Rights art. 19(2), Dec. 16, 1966, S. TREATY DOC. No. 95-20, 999 U.N.T.S. 171 [hereinafter ICCPR]; G.A. Res. 217 (III) A, Universal Declaration of Human Rights art. 19 (Dec. 10, 1948).
[2] See *id*.; Human Rights Council Res. 20/8, U.N. Doc. A/HRC/20/L.13 (June 29, 2012).
[3] ICCPR, *supra* note 1, art. 19(2).
[4] http://www.oas.org/en/iachr/expression/showarticle.asp?artID=26&lID=1
[5] See, for example, the record of Chile in the 1990s. HUMAN RIGHTS WATCH, FREEDOM OF EXPRESSION AND THE PUBLIC DEBATE IN CHILE (1998).

democracy by degrading trust in institutions, fueling political polarization, and undermining the implementation of important policies.[6] However, even if enacted in good faith, government efforts to control disinformation often overshoot their targets, repressing disinformation and legitimate expression indiscriminately. The pitfalls are plentiful: internet shutdowns to prevent the spread of fake news, criminal prosecutions based on broad categories of disfavored speech, regulatory power to remove content from social media platforms—all of these tactics have grown popular in recent years.[7] Not only do these policies violate human rights, they also create useful tools for leaders with autocratic ambitions. For established and transitional democracies alike, disinformation is a problem that defies simple solutions.

III. Democracy and Freedom of Expression in Tunisia

In 2011, the people of Tunisia put an end to decades of authoritarian rule, overthrowing President Zine al-Abidine Ben Ali with a popular uprising that inspired the wave of pro-democracy protests known as the Arab Spring.[8] The transition to democracy was not smooth. In the years following the revolution, the country endured significant periods of unrest and navigated deep tensions between Islamists and secularists that boiled over into a political crisis in 2013. After an unlikely coalition achieved a breakthrough compromise, Tunisia ratified a new constitution on January 26, 2014.[9]

The 2014 constitution laid the foundations of Tunisia's commitment to freedom of expression. Article 31 of the constitution provides, "Freedom of opinion, thought, expression, information, and publication shall be guaranteed. These freedoms shall not be subject to prior

[6] Irene Khan (Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression), *Rep. on disinformation*, ¶ 2, U.N. Doc. A/HRC/47/25 (Apr. 13, 2021).
[7] *Id.* ¶ 10–13.
[8] ALEXIS ARIEFF & CARLA E. HUMUD, CONG. RSCH. SERV., RS21666, POLITICAL TRANSITION IN TUNISIA (2015).
[9] *Tunisian National Dialogue Quartet*, ENCYCLOPAEDIA BRITANNICA (Sept. 18, 2021), https://www.britannica.com/topic/Tunisian-National-Dialogue-Quartet.

censorship."[10] Article 32 provides further protections: "The state guarantees the right to information and the right of access to information and communication networks."[11] These provisions are fully in line with international human rights standards, and they provide a strong basis for freedom of expression. But to deliver real protection, governments must match constitutional language with concrete action.

Tunisia's post-revolution record shows a genuine commitment to protecting freedom of expression, though there remain some vestiges of the Ben Ali regime's repressive practices. In general, censorship has been uncommon, independent journalism has flourished, and civil society organizations have been able to conduct advocacy on political and social issues.[12] Access to the internet is relatively widespread, and people are free to use social media platforms like Facebook and Twitter.[13] These are hard-won gains worth celebrating, but restrictions on free expression have not been fully removed from Tunisian society. Tunisia's telecommunications code allows for criminal punishment for those using communication networks to "insult or disturb others." Similarly, the penal code provides for significant prison sentences for those found guilty of publishing content "liable to cause harm to public order or public morals," and the code of military justice makes it a criminal offense to criticize the military.[14] Journalists, bloggers, and activists have faced criminal prosecution pursuant to these laws.[15] In addition to these legal restrictions, extrajudicial intimidation has contributed to a climate of fear and self-censorship, especially for women and those who speak out on controversial issues like religion and LGBTQ+ rights.[16] Such

[10] *Tunisia's Constitution of 2014*, CONSTITUTE PROJECT, https://www.constituteproject.org/constitution/Tunisia_2014.pdf.
[11] *Id.*
[12] *Freedom on the Net 2021: Tunisia*, FREEDOM HOUSE, https://freedomhouse.org/country/tunisia/freedom-net/2021 (last visited Oct. 27, 2021).
[13] *Id.*
[14] *Id.*
[15] *Id.*
[16] *Id.*

failures should serve as a reminder that despite Tunisia's great strides toward the protection of civil liberties, freedom of expression in transitional democracies tends to be fragile and incomplete.

Tunisia's political crisis of 2021 casts even more uncertainty over the country's commitment to freedom of expression. In July 2021, Tunisian president Kais Saied suspended the parliament and assumed all-encompassing governmental power, and many observers have decried these actions as amounting to a coup.[17] In the aftermath of Saied's power grab, the government has arrested journalists and shut down television stations, suggesting a troubling shift in the country's approach to civic space.[18] Although Saied has expressed his continued commitment to freedom of expression in Tunisia,[19] his actions suggest that he might harbor the types of authoritarian ambitions that rarely coincide with toleration of dissent. In this context, the government's response to a politically sensitive issue like disinformation can serve as a signal of Saied's intentions.

IV. Disinformation in Tunisia

Tunisia has not been immune from the transnational wave of disinformation that has swept through many countries with relatively free information environments. Disinformation featured prominently in the most recent parliamentary and presidential elections in 2019. It spread most virulently on Facebook, which is used by a majority of Tunisians and serves as the primary source of news for a large portion of the population.[20] In the leadup to the elections, disinformation aimed

[17] Tarek Amara & Angus McDowall, *Tunisian president moves to cement one-man rule*, REUTERS (Sep. 23, 2021, 1:34 AM), https://www.reuters.com/world/africa/tunisia-president-takes-new-powers-says-will-reform-system-2021-09-22/.
[18] *Tunisia: Freedom of Expression Violations Against Journalists and Politicians*, ARTICLE 19 (Oct. 6, 2021), https://www.article19.org/resources/tunisia-freedom-of-expression-violations-against-journalists-and-politicians/; *Tunisia shuts down TV station run by opposition party leader*, FRANCE24 (Oct. 28, 2021), https://www.france24.com/en/africa/20211028-tunisia-shuts-down-tv-station-run-by-opposition-party-leader.
[19] Vivian Lee, *Tunisia's President Holds Forth on Freedoms After Seizing Power* (Sep. 29, 2021), https://www.nytimes.com/2021/08/01/world/middleeast/tunisia-president-kais-saied.html.
[20] Mona Elswah & Philip N. Howard, *The Challenges of Monitoring Social Media in the Arab World: The Case of the 2019 Tunisian Elections*, OXFORD INTERNET INSTITUTE (Mar. 23, 2020),

at influencing voters and discrediting the process took a wide variety of forms. False reports were circulated claiming that politicians were withdrawing from the presidential race to support other candidates.[21] Inaccurate stories were spread on election day that candidates had been arrested for illegal influence campaigns and violent behavior.[22] Exit polls were falsified, and an electoral official had to go on the radio to dispel the rumor that polling station pens had been filled with erasable ink.[23] These are only some of the examples that civil society monitoring organizations were able to detect, and those organizations lacked the resources and access necessary to uncover the full picture.[24]

This proliferation of disinformation in 2019 was not random; it resulted from coordinated campaigns that were often transnational in scope. One prominent example concerned an Israeli-based firm called Archimedes Group, which targeted 13 countries and amassed a following of 2.8 million users.[25] Pages linked to this group spread disinformation around the Tunisian elections until they were discovered and removed by Facebook in May 2019, only three months before the Presidential election in September of the same year.[26] Another operation run by a Tunisia-based company was even larger, reaching approximately 3.8 million Facebook users across several African countries.[27] This campaign directed disinformation at African presidential elections including Tunisia's, and it evaded detection from Facebook until 2020, well after the 2019 election

---

https://demtech.oii.ox.ac.uk/research/posts/the-challenges-of-monitoring-social-media-in-the-arab-world-the-case-of-the-2019-tunisian-elections/#continue; Yosr Jouini, *Ahead of Tunisia elections, social media was flooded with mis- and disinformation*, GLOBAL VOICES (21 October 2019), https://globalvoices.org/2019/10/21/how-misinformation-and-disinformation-disrupted-tunisias-2019-elections/.

[21] Jouini, *supra* note 17.

[22] *Id.*

[23] Elswah & Howard, *supra* note 17.

[24] *Id.*

[25] Digital Forensic Research Lab, *Inauthentic Israeli Facebook Assets Target the World*, MEDIUM (May 17, 2019), https://medium.com/dfrlab/inauthentic-israeli-facebook-assets-target-the-world-281ad7254264.

[26] *Id.*

[27] ATLANTIC COUNCIL, OPERATION CARTHAGE: HOW A TUNISIAN COMPANY CONDUCTED INFLUENCE OPERATIONS IN AFRICAN PRESIDENTIAL ELECTIONS 2 (2020).

had ended.[28] These types of coordinated efforts can destabilize even the most established democracies, and they pose grave threats in more fragile democracies like Tunisia.

Disinformation continues to plague Tunisia even after the 2019 elections. In June and October 2020, Facebook announced that it had uncovered and removed dozens of accounts linked to coordinated inauthentic behavior that targeted users throughout African and the Middle East.[29] Misinformation about the COVID-19 pandemic has proliferated as well, undermining public health efforts in a country already straining to keep the virus under control.[30] These recent developments indicate that Tunisia's struggles with disinformation are not likely to go away on their own.

V. Options for Addressing Disinformation in Tunisia

The options for reducing the spread of disinformation in Tunisia can be grouped into two broad approaches: A) a state-centric approach, and B) a civil society-focused approach. Evaluating the merits of these approaches requires not only a narrow analysis of a strategy's effectiveness in controlling disinformation, but also a broader assessment of its potential impacts on freedom of expression and democracy in Tunisia.

A. A State-Centric Approach to Addressing Disinformation

Tunisia could seek to address disinformation by expanding the power of the state to regulate the spread of information in the country, perhaps by authorizing an agency to exercise online content controls, or by expanding the role of the judicial system in policing disinformation.

---

[28] *Id.*
[29] *May 2020 Coordinated Inauthentic Behavior Report*, FACEBOOK (June 5, 2020), https://about.fb.com/news/2020/06/may-cib-report/; *October 2020 Coordinated Inauthentic Behavior Report*, FACEBOOK (Nov. 5, 2020), https://about.fb.com/news/2020/11/october-2020-cib-report/.
[30] Fredj Zamit, Arwa Kooli & Ikram Toumi, *An examination of Tunisian fact-checking resources in the context of COVID-19*, 19 JOURNAL OF SCIENCE COMMUNICATION 1 (2020).

Tunisian officials would not have to look far to find a foreign jurisdiction on which to model such an effort—France took a state-centric approach to its own problems with disinformation in 2018. Less than a year after the disinformation-riddled Brexit campaign and U.S. presidential race of 2016, France endured its first major encounter with contemporary forms of disinformation in its 2017 presidential election. Then-candidate Emmanuel Macron became a target of several fake news stories, some of which were distributed by a coordinated Russian interference campaign.[31] Macron nevertheless won the presidency, and once in office, he sought a regulatory response to the issues of disinformation that had plagued him during the campaign. This effort resulted in French law no. 2018-1202, regarding the fight "against information manipulation".[32]

The French law expands state power to combat disinformation in both the judicial system and an administrative agency. The law introduced the opportunity for a rapid legal injunction to halt the spread of "fake news" in the leadup to an election.[33] A political candidate or party can apply to a judge for an order requiring platforms like Facebook, or internet service providers, to stop the circulation of the identified misleading information.[34] The judge must decide on the order within 48 hours.[35] The law also vests new powers in the CSA (Conseil supérieur de l'audiovisuel), France's national broadcasting agency. The CSA gained the authority to suspend the broadcast of television services controlled or influenced by foreign governments, if the agency finds that they pose a risk to certain fundamental interests.[36] The law also introduced new transparency

[31] JEAN-BATISTE JEANGÈNE VILMER, SUCCESSFULLY COUNTERING RUSSIAN ELECTORAL INTERFERENCE (CSIS, 2018); Rachael Craufurd Smith, *Fake news, French Law and democratic legitimacy: lessons for the United Kingdom*, 11 JOURNAL OF MEDIA LAW 51, 56 (2019).
[32] Loi 2018-1202 du 22 décembre 2018 relative à la lutre contre la manipulation de l'information [Law 2018-1201 of December 22, 2018 relating to the fight against information manipulation], JOURNAL OFFICIEL DE LA REPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF FRANCE], Dec. 23, 2019; *Against information manipulation*, GOUVERNEMENT.FR, https://www.gouvernement.fr/en/against-information-manipulation (last visited Oct. 27, 2021).
[33] *Against information manipulation*, *supra* note 29.
[34] Smith, *supra* note 28, at 60.
[35] *Id.*
[36] *Against information manipulation*, *supra* note 29.

requirements for online platforms.[37] The French government is confident that this law provides greater protection against disinformation without restricting freedom of expression.[38]

Apt as it may be for the French context, would a similar state-centric approach meet be advisable in the Tunisian context? It certainly has its merits. A state-centric approach like the French law may prove the most immediately effective, because it carries the force of law and provides for swift, mandatory action. This approach also leaves nothing up to chance, as it does not require the government to rely on other actors to perform the regulatory function like any civil society-focused approach would.

However, there are critical differences between the French and Tunisian contexts which render a state-centric approach far riskier in the latter. Unlike Tunisia, France is a well-established democracy with deeply embedded norms protecting freedom of expression. As discussed above, Tunisia is still shedding the remnants of repression left over from the Ben Ali regime. These differences are reflected in the reports of monitoring organizations like Freedom House, which rates France's protection of civil liberties as a 52/60 compared to Tunisia's 38/60.[39] The same organization rates internet freedom in France as a 78/100, compared to Tunisia's 63.[40] In a society like Tunisia, where the government already retains troubling amounts of power to restrict free expression, expansions of state power to regulate disinformation should be viewed with caution.[41] Such powers in the hands of would-be autocrats can serve as tools for crushing dissent and

[37] Smith, *supra* note 28, at 53.
[38] *Against information manipulation*, *supra* note 29.
[39] *Freedom in the World 2020: France*, FREEDOM HOUSE, https://freedomhouse.org/country/france/freedom-world/2020 (last visited Oct. 27, 2021); *Freedom in the World 2020: Tunisia*, FREEDOM HOUSE, https://freedomhouse.org/country/tunisia/freedom-world/2020 (last visited Oct. 27, 2021).
[40] *Freedom on the Net 2021: France*, FREEDOM HOUSE, https://freedomhouse.org/country/france/freedom-net/2021 (last visited Oct. 27, 2021); *Freedom on the Net 2021: Tunisia*, *supra* note 12.
[41] Elswah & Howard, *supra* note 17; Andreas Jungherr & Ralph Schroeder, *Disinformation and the Structural Transformations of the Public Arena: Addressing Actual Challenges to Democracy*, 7 SOCIAL MEDIA + SOCIETY 1, 5 (2021).

consolidating power—an especially relevant concern for Tunisia given the actions of President Saied in 2021.[42] Tunisian officials should carefully consider these risks before jeopardizing the country's progress on freedom of expression. Tunisian civil society and the international community should not hesitate to raise the alarm if the government begins down this route.

B. A Civil Society-Focused Approach to Addressing Disinformation

Rather than empowering government agencies to control disinformation, Tunisia could instead seek to empower civil society, including watchdog organizations, journalists, and ordinary citizens, to achieve the same goals. This approach requires a multifaceted effort. First, the government can improve its commitment to freedom of information. This includes both governmental transparency and regulation of social media to require more access to its data, which organizations can then use to track down and expose disinformation.[43] Second, the government must strengthen its protection of the public sphere. That requires not only governmental restraint in resisting the temptation to restrict free expression, but also increased governmental protection for journalists and activists who face intimidation and violence for doing their work.[44] Fourth, the government can make funding available to civil society organizations seeking to monitor and control disinformation. These organizations have thus far lacked the resources to conduct sufficient efforts.[45] Fifth, the government can improve personal data protection for individuals on social media platforms, so that their data cannot be so easily exploited for content targeting.[46] Sixth, the government can invest in the digital literacy of its population. Media information and digital literacy should be taught in schools, and the government should support similar campaigns targeted

[42] Khan, *supra* note 6, at 9–13.
[43] Khan, *supra* note 6, at 17–20; Elswah & Howard, *supra* note 17.
[44] Khan, *supra* note 6, at 18.
[45] Elswah & Howard, *supra* note 17.
[46] *Id.*

at adults.[47] With these efforts, a thriving Tunisian public sphere could emerge as a check against disinformation without any of the risks that come with a state-centric approach.

Tunisian officials need not question the capacity of civil society organizations to rise to the challenge—many have already begun responding to disinformation despite significant constraints. In the leadup to the 2019 elections, multiple civil society organizations engaged in social media monitoring, but their efforts were frustrated by funding constraints and lack of access to data from Facebook.[48] Nawaat, an organization devoted to fact-checking and accountability journalism, has experienced an increase in popularity since its founding in 2018, particularly around its COVID-19 fact-checking.[49] Most recently, the National Union of Tunisian Journalists created a disinformation monitoring unit in March 2021.[50] The work of these organizations should give assurance to the Tunisian government that its vibrant civil society can tackle the problem of disinformation if given the chance.

VI. Looking Forward

Tunisia is enduring another critical point in its post-revolution transition, one that may deepen its democracy, alter it fundamentally, or even bring it to an end. No matter what government emerges from this crisis, the problem of disinformation will continue to circulate in the public sphere. The government's response to that problem will serve as a bellwether for the direction of Tunisian democracy going forward. If the government seeks to control disinformation by tightening its grip over the public sphere, heedless of the collateral damage to freedom of expression (or worse, desiring such damage), then Tunisian citizens and the international

[47] Khan, *supra* note 6, at 18.
[48] Elswah & Howard, *supra* note 17.
[49] *Tunisia: A Growing Appetite for Fact Checks*, INTERNATIONAL MEDIA SUPPORT, https://www.mediasupport.org/tunisia-a-growing-appetite-for-fact-checks/ (last visited Oct. 27, 2021).
[50] *Freedom on the Net 2021: Tunisia*, *supra* note 12.

community may rightly worry for the broader health of Tunisia's democracy. If, however, the government empowers civil society to fight disinformation, recognizing the extraordinary capacity the Tunisian people have shown time and again, the country can simultaneously meet this challenge and demonstrate its commitment to democratic principles.

Dis-informed Democracy: Election Monitoring and the Legal Fight for Truth at the Ballot Box

I. Introduction

Elections are intrinsically tied to social movements as each phenomenon influences the outcome of the other.[1] Though the election monitoring process was not conceived with the role of disinformation in mind, it has become clear that there is a growing need to monitor and combat election disinformation. Two of the most prominent examples of recent election disinformation campaigns are the 2016 U.S. presidential election, which saw a coordinated disinformation campaign by the Internet Research Agency (a Russian troll farm),[2] and the 2020 U.S. presidential election, which led to the #StopTheSteal campaign and resulting social movement that perpetuates disinformation about the integrity of the election results. However, the United States is certainly not alone in facing this challenge, from a Russian disinformation campaign in the lead up to the Brexit referendum[3] to an Israel-based influence campaign in the 2019 Nigerian presidential elections[4] to the spread of disinformation on WhatsApp in Brazil,[5] election disinformation is a potential spoiler for elections around the globe. While some progress has been made to develop

[1] Michael T. Heaney, *Elections and Social Movements*, *in* The Wiley-Blackwell Encyclopedia of Social and Political Movements (2013).
[2] Beata Martin-Rozumilowicz & Rasto Kuzel, *Social Media, Disinformation and Electoral Integrity* (Int'l Found. for Electoral Sys., Working Paper, 2019), at 7, https://www.ifes.org/sites/default/files/ifes_working_paper_social_media_disinformation_and_electoral_integrity_august_2019_0.pdf.
[3] Agnieszka Legucka, *Russia's Long-Term Campaign of Disinformation in Europe*, Carnegie Europe (Mar. 19, 2020), https://carnegieeurope.eu/strategiceurope/81322.
[4] Isabel Debre, *Israeli disinformation campaign targeted Nigerian election*, AP News (May 17, 2019), https://apnews.com/article/8c5eec1f55bd4e209edbf5f9401e87c2.
[5] Luiza Bandeira et al., *Disinformation in Democracies: Strengthening Digital Resilience in Latin America*, Atlantic Council (Mar. 28, 2019), https://www.atlanticcouncil.org/in-depth-research-reports/report/disinformation-democracies-strengthening-digital-resilience-latin-america/.

institutions to combat the issue, such as civil society fact-checking initiatives[6] and digital ethics pacts,[7] there has been a limited shift in re-defining a role for election monitors. Yet, as "specialized human right defenders," election monitors should seek to monitor disinformation, which poses a clear threat to the right to free and fair elections.[8]

This Chapter seeks to delineate three different contexts in which election monitoring and disinformation overlap and identify potential avenues for strengthening legal protection going forward to support the right to free and fair elections. Part II gives a brief background on the framework of election monitoring. Part III then examines how specific legal frameworks that govern election monitors could be strengthened to address disinformation. Part IV illustrates how election monitoring itself can propagate disinformation and outlines the legal challenges to combating this phenomenon. Finally, Part V highlights the threat of legal action against domestic civil society monitors and explores how a rights-based approach could better protect election monitors.

[6] *Building Civil Society Capacity to Mitigate and Counter Disinformation,* Countering Disinfo, (Apr. 3, 2021), https://counteringdisinformation.org/topics/csos/2-fact-checking.
[7] *Digital Ethical Pact*, Countering Disinfo (Mar. 5, 2021), https://counteringdisinformation.org/interventions/digital-ethical-pact; Silvia Higuera, *On the initiative of journalists' association, political parties in Uruguay to sign a pact against misinformation*, LatAm Journalism Review (Apr. 24, 2019), https://latamjournalismreview.org/articles/on-the-initiative-of-journalists-association-political-parties-in-uruguay-to-sign-a-pact-against-misinformation/.
[8] Declaration of Global Principles for Nonpartisan Election Observation and Monitoring by Citizen Organizations and a Code of Conduct for Non-partisan Citizen Election Observers and Monitors, Apr. 3, 2012, at 2 [hereinafter Declaration of Global Principles], available at https://gndem.org/assets/pdfs/DoGP_en.pdf.

II. Background

Free and fair elections are an integral part of functioning democracies.[9] In the last fifty years, as the number of countries with a democratic system of governance has increased,[10] the field of election monitoring has seen exponential growth.[11] Typically, the monitoring process occurs across multiple phases before, during, and after election day, covering all aspects from voter education and registration to ballot design and security to the environment at polling stations.[12] After election day, the observing entity often puts out a preliminary statement, within forty-eight hours, of its findings on the legitimacy of the election and later submits a more detailed final report.[13] Election monitoring organizations include international organizations and domestic civil society. Crucially, election monitors are independent entities; they observe but do not intervene, and, in the case of international observers, they must be invited by the host state.[14] Election monitoring has been credited with a variety of benefits such as providing reassurance in contested elections, offering advice on how to improve electoral processes, and making incumbent turnover

[9] Document of the Copenhagen Meeting of the Conference on the Human Dimension of the CSCE, June 29, 1990, https://www.osce.org/files/f/documents/9/c/14304.pdf, ("common determination to build democratic societies based on free elections and rule of law"); African Charter on Democracy, Elections, and Governance, Jan. 30, 2007, https://au.int/sites/default/files/treaties/36384-treaty-african-charter-on-democracy-and-governance.pdf, ("Promote the holding of regular free and fair elections to institutionalize legitimate authority of representative government as well as democratic change of governments").
[10] Drew Desilver, *Despite Global Concerns About Democracy, More Than Half of Countries Are Democratic*, Pew Rsch Ctr. (May 14, 2019), https://www.pewresearch.org/fact-tank/2019/05/14/more-than-half-of-countries-are-democratic/.
[11] Patrick Merloe, *Authoritarianism Goes Global: Election Monitoring v. Disinformation*, 26 J. of Democracy 79, 79 (2015).
[12] *Manual for OAS Electoral Observation Missions*, Org. of Am. States, http://www.oas.org/es/sap/docs/manual_misiones_publicado_en.pdf; *Handbook for European Union Election Observation Missions*, Eur. External Action Serv., https://www.ec-undp-electoralassistance.org/wp-content/uploads/2018/08/undp-contents-publications-handbook-for-UE-election-observation-missions-English.pdf.
[13] *Manual for OAS Electoral Observation Missions*, *supra* note 12, at 13; *Handbook for European Union Election Observation Missions*, *supra* note 12, at 109-112.
[14] *Handbook for European Union Election Observation Missions*, *supra* note 12, at 13, 17.

more likely.[15] Arguably, election monitoring is such a diffuse practice, with eighty percent of national elections being monitored, that it is an international norm.[16]

As election monitoring has proliferated, there has been a growing recognition of the need for a common framework of the rights and responsibilities of election monitors. The two leading frameworks are codes of conduct drafted by the National Democratic Institute (NDI) and the Global Network of Domestic Election Monitors (GNDEM). In 2005, NDI drafted the Declaration of Principles for International Election Observation and the Code of Conduct for International Election Observers which has since been endorsed by fifty-five intergovernmental and international organizations.[17] In 2010, GNDEM drafted a Declaration of Global Principles for Nonpartisan Election Observation and Monitoring by Citizen Organizations and a Code of Conduct for Non-partisan Citizen Election Observers and Monitors.[18] The United Nations General Assembly has recognized the Declaration of Global Principles in three different resolutions.[19] While these documents are non-binding, they seek to define the scope of the role of electoral monitors and the modes of conduct they should undertake. Further, non-binding resolutions can be influential in creating customary international law, which is binding, by informing state practice.[20]

---

[15] Judith Kelley, *Election Monitoring: Power, Limits, Risks,* Council on Foreign Rel. (Mar. 29, 2012, 12:36pm), https://www.cfr.org/expert-brief/election-monitoring-power-limits-and-risks.
[16] Susan D. Hyde, *Catch Us If You Can: Election Monitoring and International Norm Diffusion*, 55 Am. J. of Pol. Sci. 356, 356 (2011).
[17] Declaration of Principles for International Election Observation and the Code of Conduct for International Election Observers, Oct. 27, 2005, at 1 [hereinafter Declaration of Principles for International Election Observation], available at https://www.ndi.org/sites/default/files/1923_declaration_102705_0.pdf.
[18] Declaration of Global Principles, *supra* note 8, at 1.
[19] Merloe, *supra* note 11, at 81.
[20] Avery Davis-Roberts & David J. Carroll, *Using international law to assess elections*, 17 Democratization 416, 420-421 (2010).

III. Incorporating Disinformation into Existing Frameworks

Despite the efforts to standardize election monitoring processes, there has been a lack of progress on incorporating the monitoring of disinformation into existing frameworks. While disinformation is an issue that cuts across sectors, and election monitoring alone will not stop its proliferation, it still can, and should be, incorporated into the reporting process. Currently, there is no common international framework on best practices for election monitors to track disinformation or consistent methodology to quantify the impact of that disinformation.

Neither of the Declaration of Global Principles nor codes of conduct address disinformation. While these guidelines are not intended as detailed handbooks that break down every practical detail, they offer an overview of what election monitoring organizations ideally do. For example, the Declaration of Global Principles lists twenty-six elements that may be monitored throughout the electoral process, and disinformation is not among them.[21] The list does include the monitoring of media reporting, but that is framed more narrowly in terms of media bias, such as the amount of time spent reporting on each candidate.[22]

Individual institutions are making noteworthy efforts to incorporate disinformation observation into their electoral monitoring processes. The European Union (EU) and the Organization for Security and Co-operation in Europe (OSCE) are working to develop methodology on how to observe the use of social media during elections and electoral campaigns respectively.[23] Certain organizations, such as NDI, have started sending social media analysts on their electoral observation missions.[24] While these are welcome advances, the two most common international frameworks, which seek to create international norms and inform state practice,

[21] Declaration of Global Principles, *supra* note 8, at 9-10.
[22] Declaration of Global Principles, *supra* note 8, at 9-10.
[23] Martin-Rozumilowicz & Kuzel, *supra* note 2, at 13.
[24] Martin-Rozumilowicz & Kuzel, *supra* note 2, at 13.

should be updated to incorporate the need to monitor disinformation thus providing a baseline that prioritizes the threat of disinformation.

IV. The Threat of Zombie Monitoring

Ideally, election monitoring organizations are a means to combat disinformation, but some have intentionally perpetuated disinformation themselves. "Zombie monitors" are election monitoring organizations with little transparency that rubberstamp elections in authoritarian countries as free and fair.[25] This phenomenon began in States of the former Soviet Union to combat the influence of the OSCE's international election observer missions but has since occurred in a variety of contexts.[26] Both the Russian-led Commonwealth of Independent States (CIS) and the Chinese-led Shanghai Cooperation Organization (SCO) have been accused of supporting zombie monitors and notably neither organization has signed on to the Declaration of Principles for International Election Observation.[27]

In other instances, instead of creating their own organizations, incumbent governments have sought to influence the international election observers that they invite. Given the prevalence of international election monitoring in national elections, incumbent governments in limited democracies have sought to accept international election observation. Incumbents can avoid criticism for banning observers while manipulating the observation, so as to cleanse disputed elections. These governments invite legitimate international election observers but seek to influence their results. For example, in the 2013 presidential and parliamentary elections in Zimbabwe, incumbent president Robert Mugabe warned international observers from the South

[25] Casey Michel, *The Rise of the Zombie Monitors*, The Diplomat (Apr. 30, 2015), https://thediplomat.com/2015/04/the-rise-of-the-zombie-monitors/.
[26] Rick Fawn, *Battle over the box: international election observation missions, political competition, and entrenchment in the post-Soviet space*, 82 Int'l Affairs 1133, 1133 (2006).
[27] Christopher Walker & Alexander Cooley, *Vote of the Living Dead*, Foreign Policy (Oct. 31, 2012, 4:16 PM), https://foreignpolicy.com/2013/10/31/vote-of-the-living-dead/.

African Development Community (SADC) that Zimbabwe would pull out of the community if it "decides to do stupid things."[28] SADC observers made positive statements about the elections while a domestic organization, the Zimbabwe Election Support Network, had numerous criticisms of the process.[29] Though these positive statements alone are not evidence that the SADC observers were in fact swayed by President Mugabe, it highlights the pressure that international observers, who are invited guests to the host country, face.

While there are international legal obligations to have fair and free elections,[30] election monitors themselves have no enforcement mechanisms or even legal requirements to address what they observe.[31] In the case of zombie monitoring organizations, who the incumbent government often supports, there are few options for legal recourse. One potential solution is to strengthen the norms around legitimate election monitoring, such as those laid out in the Declaration of Global Principles, to create an increasingly clear contrast between legitimate election monitors and zombie organizations.[32]

V. The Threat to Civil Society

Another option for authoritarian incumbent administrations, aside from creating a zombie monitoring organization or influencing an international observation mission, is to suppress domestic civil society organizations that do monitor elections. For example, the fairness of Azerbaijan's recent presidential elections has been questioned for numerous years and the current

[28] Merloe, *supra* note 11, at 89.
[29] Merloe, *supra* note 11, at 89.
[30] *See* International Covenant on Civil and Political Rights art. 25(b), Dec. 16, 1966, S. Exec. Rep. 102-23, 999 U.N.T.S. 171; Universal Declaration of Human Rights art. 21, G.A. Res. 217 (III) A, U.N. Doc. A/RES/217(III) (Dec. 10, 1948).
[31] Katharine A. Wagner, *Identifying and Enforcing "Back-End" Electoral Rights in International Human Rights Law*, 32 Mich. J. Int'l L. 165, 205 (2010).
[32] Max Bader, *Disinformation in Elections*, 20 Sec. and Hum. Rts. 24, 33-34 (2018).

president has held the office since 2003.[33] Anar Mammadli is chairman of the Election Monitoring and Democracy Studies Center (EMDS), an Azerbaijani civil society organization that carries out domestic election monitoring.[34] In October 2013, EMDS found election irregularities in the 2013 Azerbaijani presidential elections.[35] In December 2013, Mammadli was arrested and later convicted by a national court and sentenced to five years in prison for several offenses including illegal entrepreneurship, tax evasion, and abuse of power.[36] Mammadli brought a case before the European Court of Human Rights (ECtHR) claiming that he was illegally detained based on political motivations.[37] The Court found that there were not sufficient facts to show that Mammadli had committed the criminal offenses he was sentenced for.[38] Instead, the Court found that the Azerbaijani government had restricted Mammadli's liberty "to silence and punish" him "as a civil society activist for his activities in the area of electoral monitoring" which violated Article 18 (limitation on use of restrictions on rights) of the European Convention on Human Rights read in conjunction with Article 5 (right to liberty and security).[39] Further, the Court rejected an argument by the government that Mammadli "could not be politically motivated because he was not an opposition leader or a public official."[40]

While this ruling did offer a remedy for Mammadli's wrongful conviction, it had limited effect. The ruling was fact-dependent and focused on a specific, wrongful, criminal charges imposed under false pretenses. While rectifying this specific error is important, it offers little to

[33] Nailia Bagirova & Margarita Antidze, *Azeri president's supporters heckle as observers declare election unfair,* Reuters, (Apr. 12, 2018, 6:38 AM), https://www.reuters.com/article/us-azerbaijan-election-monitors/azeri-presidents-supporters-heckle-as-observers-declare-election-unfair-idUSKBN1HJ1GW.
[34] Mammadli v. Azerbaijan, App. No. 47145/14, ¶ 7 (July 19, 2018), http://hudoc.echr.coe.int/eng?i=001-182178.
[35] *Id.* at ¶ 11.
[36] *Id.* at ¶¶ 14, 31.
[37] *Id.* at ¶ 41.
[38] *Id.* at ¶ 63.
[39] *Id.* at ¶¶ 104-105.
[40] *Id.* at ¶103.

prevent the suppression of other civil society organizations in the country. In fact, it is unlikely that even specific cases will actually be remedied given that ninety-six percent of ECtHR cases involving Azerbaijan in the last ten years remain pending and have yet to be implemented.[41]

Given the widespread acceptance of election monitoring, one option to offer a more comprehensive remedy for civil society organizations seeking to bring cases, would be to create more explicit obligations on states to support observation mechanisms. For example, the African Charter on Democracy, Elections, and Governance, a binding treaty, contains an obligation that states "create a conducive environment for independent and impartial national monitoring or observation mechanisms."[42] Irrespective of the wrongful detention, under such an obligation, the Azerbaijani government in *Mammadli* would have been in clear of violation of an obligation to create a conducive environment.

Alternatively, as opposed to creating a stronger state obligation, there could be a reconsideration of a rights-based approach to freedom from disinformation enforced through election monitoring. Currently, most human rights discussions around election disinformation have centered on free speech rights and the need to protect them. For example, earlier this year, a Canadian law seeking to curb election misinformation was struck down by the Ontario Superior Court as a violation of the right to free speech.[43] However, as the UN Human Rights Committee has noted, the right of political participation (Article 25 of the International Covenant on Civil and Political Rights), encompasses a right of voters to form their own independent opinion, free of "compulsion, inducement or manipulative interference of any kind."[44] Thus, voters' rights are

[41] *Azerbaijan*, European Implementation Network, https://www.einnetwork.org/azerbaijan-echr.
[42] African Charter on Democracy, Elections, and Governance, Jan. 30, 2007, https://au.int/sites/default/files/treaties/36384-treaty-african-charter-on-democracy-and-governance.pdf
[43] Elizabeth Thompson, *Law prohibiting election misinformation struck down,* CBC News (Mar. 14, 2021 4:00AM), https://www.cbc.ca/news/politics/elections-misinformation-court-free-speech-1.5948463.
[44] UN Human Rights Committee, General Comment 25, 1996, point 19.

being violated by the dissemination of disinformation and, arguably, election monitoring could play a role in protecting that right.

VI. Conclusion

Election disinformation continues to be a serious threat to fair and free elections around the world. While the election monitoring process did not develop with disinformation in mind, existing frameworks should be updated to reflect the growing challenge of disinformation and the role election monitors can play. Strengthening these frameworks and the norms of legitimate election monitors can also serve as a way to draw further distinction between legitimate and zombie monitoring. Finally, individual cases have upheld the liberties of domestic organizations to conduct election monitoring, but a broader rights-based approach that emphasizes the rights of voters to form an independent opinion, and the role that election monitoring can play in enforcing that, should be established to offer more comprehensive protection.

The Multinational Disinformation Red Pill: Legal Solutions to Online Disinformation and Right-Wing Populism Across Western Europe

I. Introduction

When Trump-backed supporters raided the United States Capitol building on January 6th, 2021, heads of state across the Western European Union expressed shock and dismay.[1] However, Western Europe is no stranger to both right-wing movements and the disinformation that fuels them. In this chapter, I examine how internet disinformation has strengthened populist movements across Western Europe. I then look at the 2018 Italian General Elections as an illustrative case of this phenomenon. The next part of my paper then discusses several international legal agreements including the International Covenant on Civil and Political Rights (ICCPR) and the International Covenant on Economic, Social, and Cultural Rights (ICESCR), to determine whether these agreements have curbed or recognized the proliferation of such disinformation. It finally discusses these findings' implications for the global community.

II. Disinformation and Right-Wing Populism

The growth of right-wing populism and media disinformation presents unique challenges for EU leadership, with the percentage of European citizens voting for right-wing populist candidates rising from seven to twenty-five percent over the last century.[2] Given that right-wing populists have become an increasingly prominent issue for national governments within the EU and for the EU parliament itself, EU leadership must answer difficult questions. These questions

[1] Emma Anderson, *'This is not America': Europe reacts as Trump supporters riot in US capital,* POLITICO Jan. 6, 2001, https://www.politico.eu/article/donald-trump-us-capitol-riot-europe-reacts/.
[2] Jan Erik Grindheim, *Why Right-Leaning Populism has Grown in the Most Advanced Liberal Democracies of Europe*, in THE POLITICAL QUARTERLY, 758, 757-771 (John Wiley & Sons, Ltd., 2019).

include how to define right-wing populism, how to understand its spread, and how to determine what legal measures are available to combat its presence.

To analyze right-wing populism's impact, it is necessary to establish a working definition of both populism and right-wing populism. While legal scholars differ on how to define populism, popular discourse focuses on two of its characteristics. First, populists believe the government's role is to reflect the general will of the people, where the "people" refer to an imagined 'heartland' of 'a virtuous and unified population."[3] Second, populists believe that the current political establishment has failed to accomplish this task.[4] For right-wing populism, political scientists add an additional dimension. They write that, unlike mainstream populists, right-wing populists believe in a group of 'others' in society, such as minorities or immigrants, who do not belong to the people model.[5] Prominent political groups across Europe have been found to fit this definition, including ones such as the Five Star Movement (Italy), Alternative for Germany (Germany), Freedom Party (Austria), National Front (France), Sweden Democrats (Sweden) and the Finns Party (Finland).[6]

For each of these political groups, internet disinformation has played a role in their respective rises in popularity. Legal scholars also differ as to the exact definition of disinformation. However, the European Commission's High-Level Expert Group on Fake News and Online Disinformation (HLEG) has found the word to encompass "all forms of false, inaccurate, or

[3] Cas Mudde, *The Populist Zeitgeist*, in GOVERNMENT AND OPPOSITION, 545, 541-563 (Autumn, 2004).
[4] *Id*.
[5] DANIELE ALBERTAZZI, TWENTY-FIRST CENTURY POPULISM THE SPECTRE OF WESTERN EUROPEAN DEMOCRACY 6 (2008).
[6] Emma Anderson, *Europe and right-wing nationalism: A country-by-country guide,* BBC (Nov. 13, 2019), https://www.bbc.com/news/world-europe-36130006.

misleading information designed, presented and promoted to intentionally cause public harm or for profit."[7]

Right-wing populists present and promote their message through several mediums. The first medium, social media, is the most attractive medium for these groups,[8] with over 41.8 percent of traffic to disinformation outlets coming from social media.[9] The second medium, advertising tools, allows right-wing strategists to narrow their strategy by identifying information about individual users and third parties.[10] By doing so, political strategists look at data ranging from demographic to lifestyle information to better understand their audience.[11] They then separate audiences into different groups (such as voter status, general and income brackets) to create "macro-target" specific messages to each group individually.[12]

Both government-sponsored media and highly partisan media outlets have played a role in spreading disinformation, with a tendency to focus on sensationalist news and by the showcasing of divisive guest-speakers, amplifying these messages to millions of viewers.[13] Strategists' use of these four mediums, in tandem, have resulted in a substantial impact, in particular the success of right-wing populist parties in Italy's 2018 general elections. Lastly, these strategies are only strengthened by the third medium used by right-wing political strategists, artificial intelligence

---

[7] DIGITAL, CULTURE, MEDIA AND SPORT COMMITTEE, DISINFORMATION AND 'FAKE NEWS': INTERIM REPORT, (House of Commons) 2017-19, HC 363, at 12 (2018).
[8] Report, European Parliament's Committee on Civil Liberties, Justice and Home Affairs, Disinformation and Propaganda– Impact on the Functioning of the Rule of Law in the EU and its Member States (2019), 32, https://www.europarl.europa.eu/RegData/etudes/STUD/2019/608864/IPOL_STU(2019)608864_EN.pdf.
[9] Hunt Allcott et al, *Social Media and Fake News in the 2016 Election*, in THE JOURNAL OF ECONOMIC PERSPECTIVES, 211, 211-236 (2017).
[10] Judit Bayer et al., *Disinformation and propaganda – impact on the functioning of the rule of law in the EU and its Member States*, Pol'y Dep't for Citizens' Rights & Const. Aff., at 59 (2019).
[11] *Id*. at 33.
[12] *Id*. at 51.
[13] Alice Marwick et al., *Media, Manipulation and Disinformation Online,* DATA & SOCIETY (May 15, 2017), https://datasociety.net/library/media-manipulation-and-disinfo-online/.

bots.[14] These robots (a/k/a "bots") are programmed to share and interact with posts across social network platforms. Today, over sixty percent of all online traffic across the Internet is attributed to them.

III. Case Study on Right-Wing Populist Disinformation Campaigns: The 2018 Italian General Elections

A. Background

On March 4th, 2018, the Italian people went to the polls and placed right-wing populists in power by notably large margins. 5-Star Movement (*Movimento 5 Stelle*), the right-wing populist group defined by its anti-immigrant, nativist platform, won the greatest percentage, thirty percent, of votes.[15] The League (*Lega - ENF*), Italy's other right-wing populist party, fell close behind, capturing seventeen percent of the vote.[16] While shocking to some, the success of right-wing populists in Italy came as no surprise to others. This is due to Italy's strong distrust of the conventional news media (forty percent of the population) and their dependency on Facebook as a main media source (fifty-four percent of the population).[17] As such, many right-wing populist groups saw the forty-three million Italians who use social networks every day as an opportunity to push them to the polls.[18]

---

[14] Douglas Guilbeault et al., *Unpacking the Social Media Bot: A Typology to Guide Research and Policy*, in POLICY AND INTERNET, 255, 225-248 (2020).

[15] Report, European Commission, Understanding Citizens' Vulnerabilities to Disinformation and Data-Driven Propaganda Case Study: The 2018 Italian General Election (2019), 32, https://op.europa.eu/en/publication-detail/-/publication/3ada7fb3-7d04-11e9-9f05-01aa75ed71a1.

[16] *Id*. at 42.

[17] Report, Reuters Institute for the Study of Journalism, Reuters Institute Digital News Report 2019 (2019), 15, https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2019-06/DNR_2019_FINAL_1.pdf.

[18] Giuliano Ambrosio, *Lo stato degli Utenti Attivi al Mese sui Social Media in Italia e nel Mondo 2019,* JULIUS DESIGN (Jan. 13, 2021), https://juliusdesign.net/28700/lo-stato-degli-utenti-attivi-e-registrati-sui-social-media-in-italia-e-mondo-2015/.

B. How the 5-Star Movement and League Manipulate Disinformation

Quantitative analyses demonstrate how the 5-Star Movement and League's tactical strategies manipulate social media algorithms to push right-wing issues and positions into the national discourse. These tactics include: (1) sharing fabricated or false news stories, (2) exploiting Facebook algorithms to bring opinion pieces to the top of feeds, and (3) sharing material to stoke animosity and distrust.

[19][20] In regard to the first tactic, both the 5-Star Movement and League were able to push substantive misleading or inaccurate information with little recourse or efforts by tech companies to limit its content. To do so, both groups frequently shared two websites through various social media platforms, italia24ore.com and inews24.it, where they attempted to mislead readers by starting their domain names with forms that resembled the credible news source ilfattoquotidiano.it.[21] By-in-large, the League and 5-Star Movement effectively accomplished their goal, with both cites making the top twenty-five most viewed URLS in Italy. The fact-checking site *Pagella Politica* found that one-half of the stories shared about the referendum were fabricated.[22]

In addition, efforts to exploit Facebook algorithms found particular success. Specifically, both parties deployed a technical strategy that placed opinion pieces that contradicted news stories to appear first in the Facebook News Feed. This strategy involved ensuring posts sympathetic to

---

[19] Ernesto Dario et al., *Italian general election 2018: digital campaign strategies. Three case studies: Movimento 5 Stelle, PD and Lega*, 2nd Int'l Conf. on Advanced Rsch. Methods and Analytics, 185, 185-192 (2018).
[20] Ivana Kottasova, *Did Fake News Influence Italy's Referendum,* CNN MONEY *(*Dec. 5, 2016), https://money.cnn.com/2016/12/05/media/fake-news-italy-referendum/index.html.
[21] Report, Foundation Open Society Institute, Mapping Italian News Media Political Coverage in the Lead Up of the 2018 General Election (2018), 11, https://op.europa.eu/en/publication-detail/-/publication/3ada7fb3-7d04-11e9-9f05-01aa75ed71a1.
[22] Ivana, *supra*, note 20.

their group received a higher ratio of comments over shares.[23] It also involved ensuring posts critical of either the League and 5-Star Movement leaders Renzi and Berlusconi received a low comment-share ratio.[24] As a result, both groups dominated the social media sphere, with the League garnering the highest number of media sources, and 5-Star Movement the largest number of overall Facebook interactions.[25]

Lastly, both groups' use of hostile narratives proved their most effective misinformation strategy. Given the salience of issues such as the economy, immigration and government corruption, right-wing strategists purposely[26] spread posts meant to evoke negative reactions towards tribalistic social issues. For instance, a 2018 study by the Universitat Politecnica de Valencia found that the 5-Star Movement and League primarily published posts that recalled facts of crime or illegality which aimed to emotionally shake the voters in an effort to both stoke fear and worry, but also attack political opponents in the process.[27] On the other hand, Partito Democratico, Italy's outgoing government party, was not found to primarily rely on negative strategies at all. Consequently, the 5-Star Movement and League were able to effectively push topics sympathetic to their cause to the center of their political discourses, with immigration, welfare system failures, and elitist corruption as the three most widely covered issues in the election.[28]

IV. International Legal Frameworks to Combat Disinformation

Given the global trend towards extremist digital misinformation, policymakers and legal scholars alike have increasingly sought mechanisms to constrain it through international law. Of

[23] Dario, *supra*, at 19.
[24] *Id*. at 97.
[25] *Id*. at 9.
[26] *Id*. at 4.
[27] Dario, *supra* note 19, at page number.
[28] Giglietto, et al., *supra* note 21, at 8.

the legal solutions proposed, Article 10 of the European Convention on Human Rights has garnered particular attention. It states:

> The exercise of these freedoms, since it carries with it duties and responsibilities, may be subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society, in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary.[29]

Consequently, legal scholars suggest judicial precedent under Article 10 supports the proposition that speech freedoms are subject to limitation if they jeopardize other goals, namely national security, public safety, and crime and disorder prevention.[30] They have also held that Article 10 is in accordance with other international legal standards.[31] These standards include those put forth by Article 20 of the International Covenant on Civil and Political Rights[32] (finding hatred that constitutes incitement to discrimination, hostility or violence to be prohibited by law) and

[29] EUROPEAN CONVENTION FOR THE PROTECTION OF HUMAN RIGHTS AND FUNDAMENTAL FREEDOMS (European Court of Human Rights 1950).
[30] Report, European Parliament's Policy Department for Citizens' Rights and Constitutional Affairs, Disinformation and propaganda – impact on the functioning of the rule of law in the EU and its Member States (2019), 92, https://www.europarl.europa.eu/RegData/etudes/STUD/2019/608864/IPOL_STU(2019)608864_EN.pdf.
[31] *Id*.
[32] INTERNATIONAL COVENANT ON CIVIL AND POLITICAL RIGHTS (United Nation Human Rights High Commissioner 1966).

Article 4 of The International Convention on the Elimination of Racial Discrimination (finding dissemination of ideas based on racial superiority or hatred punishable by law).[33]

European Parliament's Policy Department for Citizens' Rights and Constitutional Affairs itself has found Article 10 to be in tandem with the Joint Declaration by the United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression, finding:

> States have some limited flexibility under international law in deciding whether or not, and if so how, to restrict freedom of expression to protect legitimate aims while respecting the standards set out above, including to reflect their own traditions, culture and values. International law also recognises that different approaches towards restrictions on freedom of expression may be justified by the very different factual situations States may face. Neither of these variations in any way undermines the principle of universality of freedom of expression and restrictions on freedom of expression should never represent an imposition by certain groups of their traditions, culture and values on others.[34]

As such, Article 10, in accordance with similar international legal standards, suggests that freedom of expression may be limited if it poses a threat to security and equality. Consequently, while Article 10 establishes that states can curb speech, it does not give much consideration to security or equity concerns, the fact that speech may be limited on the basis of falsity alone would

[33] INTERNATIONAL CONVENTION ON THE ELIMINATION OF ALL FORMS OF RACIAL DISCRIMINATION (United Nations Human Rights High Commissioner 1965).
[34] JOINT DECLARATION ON UNIVERSALITY AND THE RIGHT TO FREEDOM OF EXPRESSION (United Nations Special Rapporteur on Freedom of Opinion and Expression 2014).

likely reduce the Article's scope.[35] As such, if certain forms of speech are criminalized in accordance with human rights standards, policymakers will need to navigate a balance between curbing the spread of disinformation while also conserving free expression liberties. This would involve a narrow approach in line with free speech protections, such as enacting laws to develop artificial intelligence bots which disseminate disinformation and identify misleading sources.[36]

Another alternative discussed in Italian Parliament but stalled in partisan deadlock in response to the 2018 general elections was that state governments could implement information processing systems to mitigate "manifestly unfounded and biased news, or openly defamatory content."[37] Such systems would allow a mechanism to account for and analyze disinformation trends without directly curbing free speech. Social media platforms would consequently respond to public pressure and utilize these tools to understand how they can combat extremism and disinformation on their respective platforms. However, it is important to note that its capabilities may be limited, given bots can only identify certain types of defamatory context, but cannot do so equitably.[38] Even supposedly neutral bots could wrongly curb speech. As a consequence, California lawmakers passed Senate Bill 1001, a transparency law to protect these free speech concerns implicated in laws such as those introduced in Italy. Under S.B. 1001, lawmakers regulate bots through transparency initiatives that prohibit bots from interacting with humans.[39] The law regulates bot expressions and requires users to disclose the identities of their bots.[40] However,

---

[35] Bayer, *supra* note 8, at 22.
[36] *Id*.
[37] Sofia Verza, *European Center for Press and Media Freedom, Tackling fake news, the Italian way,* EUROPEAN CENTER FOR PRESS AND MEDIA FREEDOM, (May 22, 2018), https://www.rcmediafreedom.eu/Tools/Legal-Resources/Tackling-fake-news-the-Italian-way.
[38] Report, Random Corporation, Counter-Radicalization Bot Research: Using Social Bots to Fight Violent Extremism (2020), 27, https://www.rand.org/content/dam/rand/pubs/research_reports/RR2700/RR2705/RAND_RR2705.pdf.
[39] Kristin Houser, *Should Bots Have a Right to Free Speech? This Non-Profit Thinks So,* FUTURISM, (May 25, 2018), https://futurism.com/robots-free-speech-rights.
[40] *Id.*

while the law seeks to regulate free speech, many critics argue it actually curbs the free speech of artificial intelligence engineers and undermines the speech goals it aims to uphold.[41] As a consequence, the implementation of artificial intelligence bots to curb misinformation in countries such as Italy may be misleading.

V. Conclusion

As trends in Western Europe indicate, the rise in right-wing populism will not abate in the near future. And as the operations of the 5 Star Movement and The League in the 2018 Italian General Elections indicate, right-wing political strategists will continue to take advantage of algorithms to push emotionally charged, divisive material to widen their political base. However, as recent analysis of both the International Covenant on Civil and Political Rights, European Convention on Human Rights, United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression, and The International Convention on the Elimination of Racial Discrimination indicate, international law may provide a means to curb the widespread use of these studies. If international governing bodies are able to do so, without violating the free-speech rights of those they seek to regulate, then a legal solution to right-wing extremist disinformation campaigns is possible.

[41] *Id.*

The Far-Right in Polish Memory Wars: The Legal Battles Over Historical Truth

I. Introduction

"The slaughter of millions of defenseless people is taking place amidst a general, sinister silence…. This silence cannot be tolerated. Whatever its motives, it is depraved. Whoever is mute in the face of murder becomes the murderer's accomplice. Whoever does not condemn it permits it."[1] These words were issued by the Polish resistance fighter, Zofia Kossak-Szczucka, who was disgusted by the indifference of Europe, the world, and her Polish neighbors towards the fate of the Jewish people during the Second World War. Despite not being Jewish herself, Kossak-Szczucka observed and reported, through her underground leaflet, *Protest*, on the Nazi liquidation of the Warsaw Ghetto throughout the summer of 1942.[2] Among her other resistance activities, Kossak-Szczucka's condemnations of the inaction of the world towards Poland's Jews earned her the recognition of Yad Vashem as a Righteous Among the Nations, a title bestowed by the State of Israel to those who risked their lives to protect victims of the Holocaust.[3]

The passage quoted above indicates that what one says, or in this case does not say, about the Holocaust could have legal implications. It is an early formulation of the legal responsibility towards truth in speech about the Holocaust.[4] Even the decision by bystanders to remain silent,

---

[1] Robert Szuchta, *Against Silence and Indifference: Why I teach about the Holocaust – reflections of a teacher in* WHY SHOULD WE TEACH ABOUT THE HOLOCAUST 55, 58 (Jolanta Ambrosewicz-Jacobs & Leszek Hońdo ed., Michael Jacobs trans., 2nd ed. 2005) (quoting "Protest - odezwa conspiracynego Frontu Odrodzenia Polski" (Protest- appeal of the conspiratorial Front of the Rebirth of Poland) *quoted in* A.K. Kunnert, Polacy-Żydzi 1939-1945: wybór źródeł (Poles-Jews 1939-45: collection of sources 213 (2001).

[2] YAD VASHEM: THE WORLD HOLOCAUST REMEMBRANCE CTR., THE RIGHTEOUS AMONG THE NATIONS DATABASE, Record Details Entry Szatkowska Zofia (Kossak), https://righteous.yadvashem.org/?searchType=righteous_only&language=en&itemId=4015763&ind=NaN.

[3] *Id.*

[4] Kossak-Szczucka's words do not carry the legitimacy of a legislature or a government but are infused with a different kind of weight. *See generally* Jeffrey C. Blutinger, *Bearing Witness: Teaching the Holocaust from a*

Kossak-Szczucka asserts, can attain the status of a crime. After the War, speech and discussion concerning the Holocaust took on new but diverging meanings across Europe: both in the West, where the ritualized commemoration of the genocide came to occupy a central place in the "civil religion" of Euro-Atlantic political culture, and in the Eastern Bloc States, where Nazi occupation was swiftly replaced with Soviet imposed dictatorship and discussion of crimes specifically against Jews threatened competing state narratives of non-Jewish anti-fascist proletarian suffering and resistance.[5]

Through a case study analysis of the evolving Polish legal landscape concerning historical memory, this Chapter will examine how right-wing and nationalist popular movements in the country occupy a novel role in motivating and shaping the debate around memory laws. This Chapter reveals the difficulties involved in using law as a means to combat historical disinformation by demonstrating how cultural-political entrepreneurs on the Polish nationalist and extreme right flank are able to use legal tools to exploit concern about alleged historical disinformation and dominate the public memory space, consequently limiting certain kinds of speech about the Holocaust.[6] First, the Chapter provides an overview of the international and European law framework that dictates in what ways States are supposed to regulate understandings

Victim-Centered Perspective, 42 Hist. Tchr. 269, 270-271 (2009) (on the narrative preference for diverse non-perpetrator perspectives on the Shoah); *see also* YAD VASHEM: THE WORLD HOLOCAUST REMEMBRANCE CENTER, *supra* note 2. (Kossak-Szczucka ultimately saw the inside of Auschwitz, gaining legitimacy to speak on such matters among both Polish nationalists and other groups. She survived the camp and lived until 1968).

[5] Stefan Van der Poel, *Memory crisis: The Shoah within a collective European memory,* 49 J. EUR. STUD. 267, 272-75 (2019).

[6] It is important to note that these legal machinations are occurring against the background of an increasingly subservient judiciary, the independence of which has been attacked by the Law and Justice party since its ascension to power in 2015. The multifrontal challenges to judicial independence and the rule of law generally in Poland are beyond the scope of this paper and are too extensive to properly explain in a footnote, though suffice it to say that the procuratorial service is politically captured, judges both new and old are experiencing heightened partisan scrutiny for their rulings, and the government has succeeded in purging certain courts of judges with suspect loyalty and replacing them with lackies. For well sourced expertise on the topic, *see generally*, RULE OF LAW, ESSENTIAL READINGS, https://ruleoflaw.pl/category/essential-readings/.

of the past and how this framework incorporates certain assumptions about the position of far-right popular movements. Second, it examines how recent legislative steps by the Polish government and the mobilization of certain political constituencies have upset the framework and demonstrated the nearsightedness of the aforementioned assumptions. Third, it lays out how civil legal actions, especially those made possible by legislation on memory, speech, and history, enables far-right and nationalist figures to become enforcers of a national self-conception of innocence through aggressive litigation, which hews close to the government's populist line and even mimics its tactics.

II. Memory Laws in International and European Law

Memory laws include a wide variety of possible state actions, which can be promulgated through local or national bodies, the legislature, executive, or judiciary, with the intention of codifying, enforcing, or announcing the official position of the regime towards a particular theme or event in history. In their purest form, memory laws consist of "legislation penalizing statements about the past," but under a broader definition can also include declarative laws giving an "official assessment of historical events"; laws on "state symbols, holidays, remembrance days, and commemorative ceremonies"; "laws on the creation of museums, erection of monuments, and organization of archives"; as well as laws passed for the purpose of "lustration" that aim at "purifying public institutions from collaborators of a former regime."[7]

Memory laws, even those with pernicious speech-chilling effects, are neither a uniquely Polish nor post-Socialist phenomenon.[8] Indeed, the ill of historical denialism is recognized both in

[7] Nikolay Koposov, Memory Laws, Memory Wars: the politics of the past in Europe and Russia 6 (2018).
[8] *See, e.g.,* 2021 Tex. Educ. Code Ann. § 28.0022(4)(A)(viii) (forbidding teachers from including in civics curriculums concepts related to slavery and racism, except as "deviations from, betrayals of, or failures to live up to the authentic founding principles of the United States, which include liberty and equality").

EU and international law, and, as it concerns the latter, derives its legitimacy from foundational treaties, including the UN Declaration of Human Rights and the International Covenant on Civil and Political Rights (ICCPR). These have been interpreted to recognize a necessary balance between the rights to free expression and the rights of victim populations of genocide to not have their suffering denied.[9] Despite the apparent sanctioning of memory laws of this type under international law, the UN Human Rights Committee remains wary of their use and has gone so far as to expressly reject as incompatible with ICCPR municipal memory laws that prescribe certain historical views under penalty of law.[10] European law is similarly active on the questions relating to memory law. During the 2008 German Presidency, the European Council adopted a framework decision on "combating certain forms and expressions of racism and xenophobia by means of criminal law" that called on member states to "take necessary actions" to ensure that "publicly condoning, denying or grossly trivializing crimes of genocide, crimes against humanity and war crimes" were punishable under the law.[11]

Resisting and challenging the far-right, especially in its attempts to deny the occurrence of the Holocaust and other catastrophic humanitarian crimes, has traditionally defined the contours of the European debate about memory. The specter of the Nazis and the horrors of the Holocaust

[9] Emanuela Fronza, Memory and Punishment: historical denialism, free speech and the limits of criminal law 53 (2018) (noting that while Article 19 of the Universal Declaration of Human Rights guarantees freedom of expression and opinion, it acknowledges restrictions as necessary to guarantee the respect for the rights and freedoms of others. Also noting that ICCPR similarly provides restrictions "expressly established and deemed necessary to ensure the respect of the rights or reputation of others."). *See also* Robert Faurisson v. France, CCPR/C/58/D/550/1993*, Commc'n. No.550/1993 (*Hum. Rts. Comm. Dec. 16, 1996) (refusing to urge the repeal of the French Gayssot Law, cautioning that legitimate restrictions on freedom of speech could exist for the protection of the reputation of victim peoples and that denialism constitutes a principal vector of anti-Semitism.)

[10] U.N. Hum. Rts. Comm., Gen. Comment no. 34, U.N. Doc. CCPR/C/GC/34 ("Laws that penalize the expression of opinions about historical facts are incompatible with the obligations that the Covenant imposes on States parties in relation to the respect for freedom of opinion and expression. The Covenant does not permit general prohibition of expressions of an erroneous opinion or an incorrect interpretation of past events.").

[11] Council Framework Decision 2008/913/JHA, art. 1 (1)(c), 2008 O.J. (L. 328/55) (EU) https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32008F0913&from=EN.

dominated legal discourse concerning memory; early municipal memory laws in Italy and East Germany were effectively direct repudiations of those countries' extreme-right pasts and supplied a legal basis for bans on the positive historical evaluation of fascism.[12] Later, expressly anti-denialist laws, which became a "pan-European phenomenon," proliferated from the late 80s into the 90s, a time that corresponded with an alarming rise in popularity and activity of the increasingly transnational European far-right.[13] The European Court of Human Rights (ECtHR) jurisprudence on the validity of municipal anti-denialist memory laws demonstrates the motivating role of nationalist and far-right actors in the passage of such legislation. The two principal ECtHR cases on this matter, Garaudy v. France[14] and Perinçek v. Switzerland,[15] deal respectively with convictions under domestic laws for denial of the Holocaust and the Armenian Genocide. Through these cases, the court delineates the limits of acceptable free speech under the European Convention on Human Rights.[16]

---

[12] Koposov, *supra* note 8, at 8-10 (2018) (noting that particularly in Western Europe anti-denialist laws were intended to "create a common European memory centered on the memory of the Holocaust as a means on integrating Europe, combating racism, and averting the national and ethnic conflicts that national narratives are likely to stimulate.").

[13] Koposov, *supra* note 8, at 8; Pauline Picco, *The European Far Right*, DIGIT. ENCYCLOPEDIA OF EUROPEAN HIST., https://ehne.fr/en/encyclopedia/themes/political-europe/political-models-make-europe-modern-era/european-far-right (last visited 9/29/2021).

[14] Garaudy v. France, App. no. 65831/01, 29 (July 7, 2003), http://melaproject.org/sites/default/files/2019-01/ECtHR%2C%20Garaudy%20v.%20France%20%28Appl.%20No.%2065831%3A01%29%2C%20Decision%2C%2024%20June%202003_0.pdf (rejecting Garaudy's application and holding that the book for which he had been convicted in France had negationist purposes, content and tone and would consequently not receive the free expression protections of Article 10 of the European Convention on Human Rights as this was contrary to the convention's spirit and would contribute to the destruction of the liberties and rights that the convention was intended to guarantee).

[15] Perinçek v. Switzerland, App. no. 27510/08 (October 15, 2015) (a case in which a Turkish nationalist seeking to test a Swiss memory law on genocide referred to the Armenian Genocide as an "international lie". The court held that Article 17, which abrogates the rights under Article 10 when an applicant seeks to rely on the latter as a means to abuse Article 10 rights by inciting hatred or violence as was the case in the preceding case, did not apply; and resultantly the Swiss conviction was a violation of free speech rights).

[16] Fronza, *supra* note 10, at 96-106 (2018) (explaining the court's distinction between "clearly established facts" and facts around which there is still "ongoing debate" and the analytical framework to determine whether Article 10 protections apply or not); *See also* European Ct. of Hum. Rts., Guide on Article 17 of the European Convention on Hum. Rts. 10 (2020) https://www.echr.coe.int/Documents/Guide_Art_17_ENG.pdf.

International and European legal authorities have conceded that free speech cannot reign absolute in public discussions concerning the tragic past of Europe, and so have cautiously laid down guidelines for how national governments should approach the matter. In doing so, however, these high-minded internationalist jurists tacitly assumed that memory laws would constrain rather than amplify the activities of right-wing popular movements. The legal environment in Poland demonstrates this assumption's folly.

III. Memory Law in Poland: Legislation and Politics

The activities of Poland's ruling Law and Justice Party (PiS) in the field of memory law represent a discontinuity in the standardly conceived position of the far-right in relation to the promulgation of legally sanctioned narratives about 20th century European history. Rather than far-right groups serving as a threat against which memory laws are tools in the fight, in Poland they are among the keenest supporters of memory laws, which they characterize as a means to combat historical disinformation that unfairly casts Poland and Poles in the role of perpetrator-collaborator instead of victim-hero.

Poland's so-called "Holocaust Law,"[17] the furor surrounding its passage, and the recent civil cases brought against historians and news outlets under its provisions illustrate this shifted paradigm. In 2018 the PiS-controlled national legislature amended the law on the Institute of National Remembrance (IPN) attaching criminal penalties[18] of up to three years imprisonment for

[17] Ustawa z dnia 26 stycznia 2018 r. o zmianie ustawy o Instytucie Pamięci Narodowej (Law of 26 Jan. 2018 amending the act of the Institute of National Remembrance), Dz. U. 2018 poz. 369. (while "Holocaust Law" is the term most often used in international media and is likely the most searchable term for an English-reading audience, it should be noted that the law includes provisions pertaining not just to the Holocaust but other egregious historical crimes deriving from the wartime and post-War era, for example those committed by Ukrainian nationalists in the Volhynia Region and associated Polish retaliations. Actors on the far-right have noted the use of "Holocaust Law" as evidence of bias against Poland in the international media.

[18] Criminal penalties of the law were ultimately removed as part of a diplomatic settlement with Israel and mounting international pressure including from the United States, though civil avenues to prosecute those who mischaracterize

anyone who "publicly and contrary to fact, attributes to the Polish Nation or Polish State responsibility or co-responsibility for Nazi crimes committed by the Third German Reich" and opening up civil avenues for NGOs to sue over alleged historical mischaracterizations.[19] The law applies equally to Polish and non-Polish citizens, with no qualifying geographic limitations.[20] This is an assertion of extraterritorial jurisdiction under the "unconditional protective principle," usually invoked to safeguard a state's security or the operation of essential government functions.[21] In a press release accompanying the law's passage, the Minister of Justice, who drafted the law, offered clarification on what exactly it was that the Polish State intended to protect: "Today the Polish government took an important step… to defend our rights, defend the historical truth, and defend Poland's good name everywhere in the world."[22] While the defense of historical truth is among the stated aims of the 2018 amendment, it is complemented by an equal or greater concern for the protection of the historical record as it concerns Poland's performance during The Second World War, which is framed as a matter of sovereign national rights.[23]

---

the Polish role in World War II remain on the books. The act amending the original law was passed at breakneck speed, in the dead of night and with little of the legislative procedure that usually accompanies the creation of law by the Polish parliament. *See,* Ustawa z dnia 27 czerwca 2018 r. o zmianie ustawy o Instytucie Pamięci Narodowej (Law of 27 June 2018 amending the act of the Institute of National Remembrance), Dz. U. 2018 poz. 1277; *See also,* Paweł Sobczak, Poland backs down on Holocaust law, moves to end jail terms, REUTERS (June 27, 2018) https://www.reuters.com/article/us-israel-poland/poland-backs-down-on-holocaust-law-moves-to-end-jail-terms-idUSKBN1JN0SD.

[19] Ustawa z dnia 26 stycznia 2018 r. o zmianie ustawy o Instytucie Pamięci Narodowej (Law of 26 Jan. 2018 amending the act of the Institute of National Remembrance), Dz. U. 2018 poz. 369.

[20] *Id.*

[21] *Id. See also* Tomasz Tadeusz Koncewicz, *On the Politics of Resentment, Mis-memory and Constitutional Fidelity: The Demise of the Overlapping Polish Consensus in* L. AND MEMORY TOWARDS LEGAL GOVERNANCE HIST. 263, 269 (Uladzislau Belavusau & Aleksandra Gliszczyńska-Grabias eds.).

[22] Koncewicz, supra note 21.

[23] For another demonstration of how sovereign national rights and history are understood as related in Poland, *see also* Stephen Mulvey, *Poles in War of Words Over Voting,* BBC NEWS (June 21, 2007) http://news.bbc.co.uk/2/hi/europe/6227834.stm (in which Jarosław Kaczyński, then-PM and current leader of the Law and Justice Party widely accepted as the true power behind the party's success, claimed that Poland's WWII dead should be given consideration in determining voting power at the European Union).

Supporters of PiS's historical policies insist that the 2018 law's adoption was intended to specifically counter the phrase "Polish Death Camps," which they contend misattributes the operation of the Holocaust on the territory of occupied Poland to the Poles themselves instead of Nazi German occupiers.[24] Scholars generally agree that this term is misleading.[25] However, prior to the passage of the 2018 law, civil measures already provided plausible avenues to recourse for those who alleged harm resulting from the use of the "Polish Death Camps" phrase and similar misrepresentations of Poland's role in the Holocaust.[26] Moreover, the Polish legal code contained criminal provisions adequate to combat historical disinformation of this kind: "Art. 55 of the Law on IPN (denying Nazi crimes); Art. 133 of the Polish Criminal Code (publicly insulting the Polish nation or Poland); Art. 212 of the Polish Criminal Code (slander); [and] Art. 216 of the Polish Criminal Code (insult)."[27]

These pre-existing provisions, which reveal a duplication of efforts toward the 2018 law's stated aims, suggest the existence of a political factor to PiS' historical policies. Indeed, while President Duda hesitated in affixing his signature to the bill in the face of international criticism, that political factor mobilized in support of the new measure, revealing themselves in public protest as representatives of the anti-Semitic far-right.[28] "Remove your yarmulke and sign the bill,"

[24] Uladzislau Belavusau, *The Rise of Memory Laws in Poland,* 29 Sec. and Hum. Rts. 36; *See also* Marc Santora, *Poland's "Death Camp" Law Tears at Shared Bonds of Suffering with Jews*, N.Y. TIMES (Feb. 6, 2018) https://www.nytimes.com/2018/02/06/world/europe/poland-death-camp-law.html (last visited 9/29/2021).
[25] Santora, supra note 24.
[26] Koncewicz, *On the Politics of Resentment* 272 n. 27 (noting three cases Osewski v Die Welt/Axel Springer, Zapaśnik v Focus Online – Tomorrow/Focus Media GMBH, and Tendera v ZDF, in which users of the phrase Polish death camps or similar formulations understood as misattributing blame for the Holocaust to Poland or Poles were taken to court as defendants prior to the 2018 law).
[27] *Id.* at 270.
[28] During PiS' first term in government, when the 2018 amendment was made, an extreme right-wing coalition party was widely understood as a credible threat to the survival of the governing majority as only a small diversion of support from PiS to their rightist opponents could have been sufficient to allow the centrist-liberals back into the driving seat. *See, e.g.,* Zosia Wanat, *Poland's Ruling Party Faces New Far-Right Election Threat*, POLITICO, (May 26, 2019), https://www.politico.eu/article/polands-ruling-party-faces-new-far-right-election-threat/.

"Defend the Truth," and "#StopAntiPolonism" read the banners of nationalist demonstrators in Warsaw demanding the President sign the bill into law.[29] Reviewing the effect of the law retrospectively, the Polish League Against Defamation (Polish: *Fundacja Reduta Dobrego Imienia – Polska Liga Przeciw Zniesławieniom*, henceforth RDI), a right-wing nationalist NGO dedicated to protecting the reputation of Polish nation from alleged misrepresentations in international media, also signaled its support for the 2018 law. RDI claimed that the amendment was both justified and effective despite the global outcry and accusations of Holocaust denial, which they suggested emerged by means of a coordinated global effort by the international media to defeat the measure, not so subtly echoing an anti-Jewish dog-whistle in the same breath as refuting accusations of anti-Semitism.[30] While memory laws, especially those with anti-denialist provisions, have historically acted to constrain the activity of far-right groups by casting their hateful rhetoric as exterior to the realm of legally protected speech, the 2018 Amendment to the Act on the IPN departs from this model. It instead dares anyone to challenge the historical innocence of Poland that is close to the hearts of the nation's far-right. Recognizing this fact, this unsavory political constituency rallied to the bill's support.

---

[29] *"Zdejmij jarmułkę, podpisz ustawę". Manifestacja narodowców przed Pałacem Prezydenckim* ("Remove your yarmulka and sign the bill". A demonstration of nationalists in front of the Presidential Palace), POLSAT NEWS (Feb. 5, 2018), https://www.polsatnews.pl/wiadomosc/2018-02-05/zdejmij-jarmulke-podpisz-ustawe-manifestacja-narodowcow-przed-palacem-prezydenckim/(noting that the protest was organized and attended by representatives of Stowarzyszenie Marsz Niepodległości (The Independence March Association), a right-wing pressure group that among other activities convenes the annual independence march in Warsaw that in the past has been well attended by representatives of Europe's neo-Nazi and extreme-right parties).

[30] Annual Report, RDI, Raport z Działności Reduty Dobrego Imienia za 2018 Rok (Report of the Activities of RDI for the Year 2018) 5-6 (2018) http://www.anti-defamation.pl/rdiplad/wp-content/uploads/2019/03/Raport-z-dzia%C5%82alno%C5%9Bci-Reduty-Dobrego-Imienia-za-2018-rok.pdf (the report stated that the international media response to the amendment was "essentially homogenous" and arose as if triggered by the "push of a button" (Polish: "w zasadzie jednorodny i uruchomiony jakby za 'przyciśnięciem guzika'")).

IV. Right-Wing Litigators

The far-right movement in Poland has not limited itself to the position of outside supporters of this new type of memory law. Perhaps alarmed by the government's hesitancy to enact the original 2018 amendment in the face of international pressure and its readiness to compromise on the bill's most punitive aspects, groups and individuals associated with the far-right have taken an active role to leverage remaining provisions in the bill and other legal tools to advance a memory campaign against alleged historical disinformation. Just two months after the amendment to the act on the IPN passed the Polish parliament, RDI initiated a lawsuit against Argentinian newspaper *Pagina/12* under article 53o of the law, which created a cause of action for NGOs to protect the reputation of the nation from slander with damages going to the state treasury.[31] In the suit, RDI demanded an apology from the newspaper for publishing an article discussing the mass murder of Jews by Poles in the village of Jedwabne during the Nazi occupation. The complaint did not allege that the massacre never took place, but instead griped that the image accompanying the article, which showed a group of Polish resistance fighters who were killed themselves after the war for continuing armed struggle against Soviet occupation, misleadingly associated patriotic Polish partisans with the massacre in Jedwabne.[32] RDI claimed that the use of the photo in the article "damages the good name of the Polish Nation (and of each Pole individually, in particular the families of victims)."[33]

In a similar case that originated in 2019, RDI covered "all legal and administrative costs" for Filomena Leszczyńska who acted as plaintiff against defendants Barbara Engelking and Jan

[31] Press Release, RDI, Statement of the Polish League Against Defamation Concerning the Claim Brought in Against Pagina12 (Mar. 5, 2018) http://www.anti-defamation.pl/rdiplad/wp-content/uploads/2018/03/2018-03-06-EN-Statement-of-the-RDI-the-claim-brought-in-against-Pagina12.pdf.
[32] *Id.*
[33] *Id.*

Grabowski, two prominent Holocaust scholars who collaborated on a book, *Night Without End: The Fate of Jews in Selected Counties of Occupied Poland*.[34] In the suit, Leszczyńska and her supporters at RDI allege that one paragraph of the 1,600-page two-volume academic tome is libelous towards her long-deceased uncle Edward Malinowski, who it erroneously describes stealing from a Jewish woman he had previously rescued and being involved in the handing over of other Jews to the German authorities.[35] Leszczyńska claimed damages of 100,000 Polish Zlotys (approximately $26,000 at the time[36]) for the harm done to her uncle's honor by the conflation of his biography with that of an identically named man, who the scholars admit was the true collaborator.[37] The judge at the district level denied the claim for damages, but still ruled in favor of Leszczyńska, ordering Engelking and Grabowski to apologize and alter future editions to reflect their mistake.[38]

While the Engelking-Grabowski-Leszczyńska and the *Pagina/12* cases may appear as two isolated examples, RDI's litigious strategy mirrors similar efforts by PiS to strangle unfavorable speech through SLAPP (strategic litigation against public participation) suits against critics. Wojciech Sadurski, a law professor active in Warsaw and Sydney, was the target of one such suit

[34] Press Release, RDI, Judge Hears Opening Arguments Today in the Libel Case Against Authors of Night Without anEnd. Fate of Jews in Selected Counties of Occupied Poland 2 (Oct. 29, 2019) http://www.anti-defamation.pl/rdiplad/wp-content/uploads/2016/10/2019-10-29-Judge-hears-opening-arguments-today-in-the-libel-case-against-authors-of-Night-Without-an-End.-Fate-of-Jews-in-Selected-Counties-of-Occupied-Poland.pdf.
[35] Agnieszka Wądołowska, *Holocaust Scholars Lose Polish Libel Case and Must Apologise for Inaccuracies in Book,* NOTES FROM POL. (Feb. 9, 2021) https://notesfrompoland.com/2021/02/09/holocaust-scholars-lose-polish-libel-case-and-must-apologise-for-inaccuracies-in-book/.
[36] Fx-exchange.com, Polish Zloty (PLN) To US Dollar (USD) on Dec. 19, 2019 https://www.fx-exchange.com/pln/usd-2019_12_19-exchange-rates-history.html (last viewed 9/30/2021) (showing that the rate of exchange a matter of months after the claim was made was 0.26095 USD per 1 PLN; 100,000 multiplied by 0.26095 equals 26,000 to two significant figures).
[37] Wądołowska, supra note 36.
[38] *Id. See further* Daniel Tilles, *Holocaust Scholars Win Appeal Against Polish Court Ordering Them to Apologise,* NOTES FROM POL. (Aug. 16, 2021) https://notesfrompoland.com/2021/08/16/holocaust-scholars-win-appeal-against-polish-court-ruling-ordering-them-to-apologise/ (reporting how Engelking and Grabowski succeeded in having the initial verdict overturned on appeal with the judge calling the lower court's decision "an unacceptable interference in the freedom of academic research and freedom of expression," though RDI has vowed to take the case up to the next and final level of judicial review).

brought by PiS.[39] The alleged offense? Sadurski had posted a tweet referring to PiS as an "organized criminal group" and advising his followers to avoid the 2018 Independence March, which he called "a parade of defenders of the white race, who have hidden for a moment their 'falangas' [a neo-Nazi symbol] and swastikas."[40] The subject of the tweet, the aggressive legal response from PiS, and the familiar anti-speech strategy of the litigation suggests a coalescence of objectives and strategies between the Polish government itself and nationalist and right-wing groups suing to defend against critics and alleged disinformation.

V. Conclusion

The involvement of Poland's domestic far-right movement in the passage and instrumentalization of the 2018 law on what one can legally say about the Holocaust demonstrates the weakness of assumptions contained within the existing international framework laid out to determine how States may regulate speech concerning the historically important crimes of genocide, war crimes, and crimes against humanity. The memory issue in Poland is part of a larger political issue. The far-right are enjoying a moment of closeness to actual power not witnessed since the destruction of the Second Republic and the start of the Second World War. This closeness has benefited the far-right as a social movement in terms of appointments, media access and state funds.[41] Additionally, however, political allegiance has been leveraged into legal changes. By the

[39] Poland: Ruling Law and Justice Party and Public Broadcaster TVP Must Drop SLAPP Defamation Lawsuits Against Law Professor Sadurski, Article 19 (Nov. 25, 2019) https://www.article19.org/resources/poland-ruling-law-and-justice-party-and-public-broadcaster-tvp-must-drop-slapp-defamation-lawsuits-against-law-professor-sadurski/.
[40] *Id. See also* Agnieszka Kublik, *"Goebbelsowskie media" nie zniesławiają TVP. Jest Wyrok w Sprawie Prof. Sadurskiego* ("Goebbelsian Media" Does Not Defame TVP. This is the Verdict in the Case of Professor Sadurski), Gazeta Wyborcza (Sept. 23, 2021) https://wyborcza.pl/7,75398,27605215,czy-o-tvp-jacka-kurskiego-mozna-twitowac-goebbelsowskie-media.html (reporting in the centrist-liberal paper in which Sadurski is found not guilty in a related case of criminal defamation under Article 212 of the Criminal Code for referring to state media TVP as a "Goebbelsian" propaganda factor).
[41] Government favor has manifested not just in the form of memory law legislation, but also cash and appointments. *See, e.g.*, Jan Kunert, *"Internetowa Husaria" za 231 tys. zł. Sprawdzamy Efekty Projektu z Pieniędzy MSZ* ("Internet Hussars" paid 231,000 PLN We Are Checking the Effects of the Project Funded by The Ministry of Foreign Affairs) KONKRET 24 (Apr. 30, 2019) https://konkret24.tvn24.pl/polska,108/internetowa-husaria-za-231-tys-zl-sprawdzamy-

passage of the 2018 law on the IPN and the erosion of judicial independence,[42] the Polish government has succeeded in signalling in law a narrative about historical events and associated complicity that have become central to Poland's socially created national identity. Unable to maintain the harshest provisions of criminality, the government has colluded to outsource the policing of this narrative to particularly litigious far-right movements eager to use civil law remedies to vindicate their "truth" over perceived historical disinformation.

efekty-projektu-z-pieniedzy-msz,920047.html (describing the use of 231,000 PLN of Ministry of Foreign Affairs awarded money by far-right internet "hussars"), *see also*, Vanessa Gera Polish State Historian Resigns After Far-right Past Revealed (describing the appoint and delayed removal of a PiS appointed Director of the Wrocław branch of the IPN credibly accused of neo-Nazi sympathies. The position has authority to prosecute and lustrate Nazi collaborators and those complicit with the crimes of the Soviet-imposed Communist regime. This director was photographed performing a Nazi salute.).

[42] *See, supra* note 6.

Disinformation and Human Rights

Authoritarian governments increasingly use the regulation of disinformation as a justification to enact laws that threaten human rights. Purported attempts to criminalize disinformation have not had the desired effect, but instead have led to the suppression of human rights. Conversely the suppression of freedoms of speech, expression, and press are often tied to the greater spread of disinformation. Without necessary checks, such as free and independent journalism or regulation of social media applications, a vacuum is created for disinformation.

There is increasing recognition and acceptance that new technology, such as deepfakes or encrypted messaging apps, require a reconsideration of human rights law. This Section highlights the shortcomings of existing international law. Few, if any, governments have been held accountable for oppressive laws, propagated in the name of preventing disinformation, and there is continuing debate concerning the need for technology companies to acknowledge their role.

Encompassing case studies in East and Southeast Asia, Africa, and Central America, this Section examines the legal regimes governing human rights concerns, such as internet blackouts. Though there is a clear need to legislate against disinformation, it must be done with clear limits, so as not to intentionally, or unintentionally, suppress human rights. Many of the case studies included, point to international law as a helpful framework with which to address oppressive domestic laws.

Disinformation and Messaging Apps in Latin America

In volatile and repressive political climates, encrypted messaging apps have emerged as a critical tool for social movements, facilitating participants' enjoyment of their rights to freedom of opinion, expression, and assembly both online and in person. Yet at the same time, messaging applications, like other social media platforms, serve as incubators for misinformation and disinformation that threaten to undermine those same rights.

This Chapter will explore how non-public social media, namely encrypted messaging platforms, fit into the disinformation "ecosystem" and human rights framework in Latin America in light of their instrumentality for social movements.[1] Part I explores the unique challenges in stemming disinformation presented by encrypted messaging apps, as well as trends in the "information ecosystem" across Latin America in which messaging apps feature prominently.[2] Part II surveys the international and regional law on human rights pertinent to addressing the disinformation pandemic. Lastly, Part III considers Nicaragua's Cybercrime Bill in light of the State's international and regional legal obligations and within the context of messaging apps' critical role in efforts to both preserve civil liberties and spread disinformation since the popular uprisings in April 2018.

[1] *See* Luiza Bandeira, et al., *Disinformation in Democracies: Strengthening Digital Resilience in Latin America*, ATLANTIC COUNCIL 8 (2019), https://www.atlanticcouncil.org/in-depth-research-reports/report/disinformation-democracies-strengthening-digital-resilience-latin-america/ ("information ecosystem").
[2] *See id.* ("information ecosystem").

I. The Unique Challenges of Closed Social Networks for Monitoring Disinformation in Social Movements

Discourse around viral disinformation often centers around Facebook, Twitter, and YouTube as the most popular social media platforms, and most commonly perceived disinformation vectors, in the United States.[3] Encrypted messaging apps like WhatsApp, Telegram, and Signal largely remain peripheral players in the U.S. social media context, although encrypted messaging has jumped in popularity since the 2020 Black Lives Matters protests and the January 2021 Capitol Riot.[4] Globally, however, encrypted messaging apps outrank other types of social media networks and serve as both potent vectors for disinformation and key services for preserving freedom of opinion and expression.[5] As the third-most popular social network globally with over two billion users worldwide, WhatsApp is the market leader amongst messaging apps.[6]

---

[3] *See Most Popular Mobile Social Networking Apps in the United States as of September 2019, by Monthly Users (in millions)*, STATISTA https://www.statista.com/statistics/248074/most-popular-us-social-networking-apps-ranked-by-audience/ (last visited Sept. 18, 2021) [hereinafter *Most Popular Mobile Social Networking Apps*] (listing Facebook as the top social media application in the United States with 169.76 million monthly users, followed by Instagram and Twitter); Brooke Auxier & Monica Anderson, *Social Media Use in 2021*, PEW RESEARCH CENTER (Apr. 7, 2021), https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ (citing 81 percent of Americans as reporting having used YouTube and 69 percent as having used Facebook).

[4] *See Most Popular Mobile Social Networking Apps*, *supra* note 3 (WhatsApp has 25.58 monthly users in the United States, followed by Telegram with 3.47 monthly users); Jason Aten, *How Signal Became the Most Popular App in the World Overnight, and Why it Matters*, INC. (Jan. 16, 2021), https://www.inc.com/jason-aten/how-signal-became-most-popular-app-in-world-overnight-why-it-matters.html (Signal became the most downloaded free app in January 2021); Ryan Gallagher, *Signal Jabs at Facebook and Navigates Growing Pains as Popularity Surges*, BLOOMBERG (May 28, 2021, 6:21 AM), https://www.bloomberg.com/news/articles/2021-05-28/signal-app-is-surging-in-popularity-and-hitting-growing-pains (Signal surpassed 100 million downloads from the App Store and Google Play by May 2021); Jacob Gursky & Samuel Woolley, *Countering Disinformation and Protecting Democratic Communication on Encrypted Messaging Applications*, BROOKINGS, (June 2021), at 4, https://www.brookings.edu/wp-content/uploads/2021/06/FP_20210611_encryption_gursky_woolley.pdf (stating Telegram claimed over 500 million users in January 2021).

[5] *See* Tess K. Bishop, *How Do You Solve a Problem Like WhatsApp? The Complicated Role of Messaging Apps in the Fight Against Disinformation and for Free Speech*, UNIVERSAL RIGHTS GROUP NYC (Apr. 2, 2021), https://www.universal-rights.org/universal-rights-group-nyc-2/how-do-you-solve-a-problem-like-whatsapp-the-complicated-role-of-messaging-apps-in-the-fight-against-disinformation-and-for-free-speech/.

[6] *See About WhatsApp*, https://www.whatsapp.com/about (last visited Sept. 18, 2021) (2 billion users); Bishop, *supra* note 5 ("In terms of total number of users, WhatsApp is almost as big as Facebook and YouTube") ("WhatsApp is the most popular messaging platform worldwide").

Closed social networks like encrypted messaging apps present a host of unique challenges in combatting the spread of disinformation. First, it is more difficult for researchers to understand what disinformation is being shared on the platforms and who originated the information.[7] WhatsApp and competitors Telegram and Signal use end-to-end encryption, a feature that safeguards users' communications from "surveillance and interception."[8] Because disinformation spreads within private channels inaccessible to the general public, it is difficult for researchers to directly analyze the content.[9] Second, users typically communicate with people they know through messaging apps, and the disinformation they receive is sent to them by people they trust.[10] This "social capital" lends credibility to, and can be exploited by, disinformation campaigns.[11] Third, public groups and channels within the apps allow for the creation of large chat rooms, which aid swift dissemination of false or misleading information that ripples out through message forwarding.[12] Finally, and perhaps most difficult, is the countervailing consideration of the applications' legitimate, and crucial, uses by human rights defenders, journalists, and political activists to communicate and coordinate democratic activities beyond the reach of government

[7] *See* Bandeira, *supra* note 1, at 12.
[8] *See* Bishop, *supra* note 5.
[9] *See* Gursky & Woolley, *supra* note 4, at 5-6 (noting that researchers can glean insight into information spreading on WhatsApp through monitoring public groups which often aim to "amplify" disinformation originated in smaller encrypted chats).
[10] *See* Elizabeth Dwoskin & Annie Gowen, *On WhatsApp, Fake News Is Fast and Can Be Fatal*, WASHINGTON POST (July 23, 2018), https://www.washingtonpost.com/business/economy/on-whatsapp-fake-news-is-fast--and-can-be-fatal/2018/07/23/a2dd7112-8ebf-11e8-bcd5-9d911c784c38_story.html; Bandeira, *supra* note 1, at 12; Gursky & Woolley, *supra* note 4, at 2, 5 (encrypted messaging apps foster a sense of trust and security amongst users which can be exploited by disinformation campaigns).
[11] *See* Gursky & Woolley, *supra* note 4, at 2.
[12] *See* Brian Chen & Kevin Roose, *Are Private Messaging Apps the Next Misinformation Hot Spot?*, N.Y. TIMES (Feb. 3, 2021), https://www.nytimes.com/2021/02/03/technology/personaltech/telegram-signal-misinformation.html (Telegram group chats can admit up to 200,000 users) (Signal group chats permit up to 1,000 people); Bandeira, *supra* note 1, at 12 (WhatsApp groups permit up to 256 users); *id.* at 13 (stating public groups on WhatsApp can facilitate "rippling distribution and amplification" of disinformation to users' smaller, "personal networks").

surveillance.[13] This privacy concern counsels against "backdoor" solutions that allow law enforcement access to chats and the disinformation therein.[14]

A. The Latin American Context

Latin American countries make up some of WhatsApp's largest user bases, with Brazil and Mexico contributing 108.4 and 62.3 million users respectively.[15] Approximately 55 percent of the population in Latin America accessed the internet with a mobile device in 2019.[16] Of total internet users between 16 and 64 years old, 93 percent of Argentines, 92 percent of Colombians, and 91 percent of Brazilians used WhatsApp monthly.[17] In those countries, 36 percent, 45 percent, and 43 percent of interviewees, respectively, reported using WhatsApp as a source of news, as did 35 percent in Mexico.[18]

Integration of WhatsApp into telecommunications carriers' "zero-rating policies" has been one driver of the app's popularity in the region.[19] Zero-rating policies allow users to access designated applications without using their mobile data allocation.[20] WhatsApp, along with Facebook and Twitter, are the most commonly included platforms.[21] Although the relative affordability of mobile devices and limited internet access through zero-rating can serve as an "on

---

[13] *See* Gursky & Woolley, *supra* note 4, at 2, 4.
[14] *See id.* at 3.
[15] *See* Harry Rollason, *What Countries are the Biggest WhatsApp Users?*, CONVERSOCIAL (July 8, 2021), https://www.conversocial.com/blog/what-countries-are-the-biggest-whatsapp-users.
[16] *See The Mobile Economy: Latin America 2020*, GSM ASSOCIATION, https://www.gsma.com/mobileeconomy/wp-content/uploads/2020/12/GSMA_MobileEconomy2020_LATAM_Eng.pdf.
[17] *See* Rollason, *supra* note 15.
[18] *See* Nic Newman, et. al, *Digital News Report 2021* 115, 123, 117, 125, REUTERS INSTITUTE, https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2021-06/Digital_News_Report_2021_FINAL.pdf.
[19] Telecommunications companies in Brazil, Colombia, Chile, Costa Rica, the Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Jamaica, Mexico, Nicaragua, Paraguay, Peru, Trinidad, and Tobago have implemented zero-rating policies. *See* Helani Galpaya, *Zero-rating in Emerging Economies*, GLOBAL COMMISSION ON INTERNET GOVERNANCE, Feb. 2017, at 2-3 (Table 1: Number of Operators Offering Zero-rated Content); *id.* at 4 (noting 15 of the 19 Latin American countries researched had at least one zero-rating program).
[20] *See* Bandeira, *supra* note 1, at 11.
[21] *See* Galpaya, *supra* note 19, at 2-3 ("Recent research from Latin America showed that among 15 countries that offered some kind of zero-rated content, 14 offered zero-rated WhatsApp or Facebook…").

ramp" to full internet usage, it remains unclear to what extent users who start with zero-rating plans progress to full internet usage.[22]

While some hail such policies as a boon to global internet access rights, others criticize their preferential treatment of some websites as an affront to net neutrality and warn of the dangers presented by bounded access to the internet.[23] One such danger is the "walled garden effect" where users may stay within bounds of the zero-rated platform instead of progressing to using the internet at large.[24] For example, users may search for information only within Facebook, lacking the research skills or financial resources to conduct broader internet searches for the information they want.[25] This creates a ripened environment for disinformation to spread unchecked. For example, researchers found that zero-rating WhatsApp in Brazil incentivized the sharing of images and memes and disincentivized sharing and opening links in messages because doing so would count towards users' data caps.[26] In sum, zero-rating plans can help users get online, but as a result users may be exposed to (dis)information they lack the financial resources or media literacy to verify.

In addition to the financial constraints that drive users towards zero-rating policies, Latin American users largely prefer to access news online, including through social media, than through traditional print, television, and radio sources.[27] Even where interest in online news is strong, few

---

[22] *See* Bandiera, *supra* note 1, at 8 ("Notably, mobile-only access was more prevalent among the country's poorer population"); Galpaya, *supra* note 19, at 8–9.

[23] *See* Corynne McSherry, et. al, *Zero Rating: What It Is and Why You Should Care*, ELECTRONIC FRONTIER FOUNDATION (Feb. 18, 2016), https://www.eff.org/deeplinks/2016/02/zero-rating-what-it-is-why-you-should-care (at Zero Rating Is Usually Only a Temporary Fix); Galpaya, *supra* note 19, at 1 ("net neutrality… [is] the concept that all content should be treated equally on the Internet.").

[24] *See* McSherry, *supra* note 23 at 3 (Walled Garden Effect); Galpaya, *supra* note 19, at 1 ("According to Facebook's data, about 50 percent of consumers who start out on Free Basics do end up paying for a data plan within a month. … [I]t is unclear if these consumers then venture on to explore content outside of Free Basics, or keep consuming the content inside Free Basics/full-version Facebook…").

[25] *See* Galpaya, *supra* note 19, at 10 (citing an example of many users in Myanmar only conducting searches through Facebook or Facebook Flex).

[26] *See* Bandeira, *supra* note 1, at 12.

[27] *See* Carolina de Assis, *Use of Instagram and WhatsApp for Online News Consumption Grows in Argentina, Brazil, Chile and Mexico: Reuters Institute*, KNIGHT CENTER FOR JOURNALISM IN THE AMERICAS (June 21, 2019), https://latamjournalismreview.org/articles/use-of-instagram-and-whatsapp-for-online-news-consumption-grows-in-

are willing to pay for it. For example, in Peru, of the 85 percent of interviewees who reported accessing news online (including through social media), only 16 percent paid for it.[28] Subscription rates are similarly low across other Latin American countries surveyed, with 18 percent paying for online news in Mexico, 17 percent in Brazil and 15 percent in Argentina.[29] Additionally, professional journalism outlets have increasingly erected paywalls to replace falling print revenues.[30] With professionally produced news available at a higher premium, "a gap between where news is now published and where information is consumed" has emerged.[31]

Together, these access phenomena contribute to the "information ecosystem" of many Latin American countries; in other words, how and where individuals access information.[32] WhatsApp in Latin America has become both a platform where disinformation proliferates and a critical tool for human rights defenders, political dissidents, and members of anti-government social movements to organize and express themselves.[33]

II. The International and Regional Human Rights Framework

Since the onset of the COVID-19 pandemic, legislatures across Latin America have introduced bills purporting to tackle health- and election-related "fake news," but many stalled amidst criticism that they pose direct threats to press freedom or enable targeting of political dissent.[34] This Part evaluates the human rights law framework set forth by the Inter-American

---

argentina-brazil-chile-and-mexico-reuters-institute/ ("Few paying for online news" section); Newman, *supra* note 18, at 115–127 (percentage paying for online news and sources of news graphics).

[28] *See* Newman, *supra* note 18, at 127.

[29] *See id.* at 125, 117, 115 (paying for online news in Brazil, and Argentina, respectively).

[30] *See* Bandeira, *supra* note 1, at 8; Newman, *supra* note 18, at 6.

[31] *See* Bandeira, *supra* note 1, at 8.

[32] *See id.*

[33] *See* Bishop, *supra* note 5; Alvaro Marañon, *How Have Information Operations Affected the Integrity of Democratic Elections in Latin America?*, LAWFARE (May 28, 2021), https://www.lawfareblog.com/how-have-information-operations-affected-integrity-democratic-elections-latin-america.

[34] *See* Júlio Lubianco, *11 Laws and Bills against Disinformation in Latin America Carry Fines, Prison and Censorship*, KNIGHT CENTER LATAM JOURNALISM REVIEW (Dec. 16, 2020),

Commission and the larger international community against which such disinformation laws must be measured.[35]

A. International Human Rights Standards

Almost all Latin American countries have ratified the International Covenant on Civil and Political Rights (ICCPR), and parties to the treaty incur an obligation under international law to give effect to the rights contained within it.[36] Article 19 of the ICCPR recognizes two distinct, but intertwined, rights: "the right to hold opinions without interference" and "the right to freedom of expression."[37] According to the United Nations Human Rights Council (UNHRC), freedom of opinion "comprises two dimensions: an internal dimension closely connected to the right to privacy and freedom of thought and an external dimension related to freedom of expression."[38] Freedom of expression, on the other hand, encompasses "freedom to seek, receive, and impart information" across borders and in any form.[39]

"While freedom of opinion is absolute," freedom of expression can be limited by a state if they satisfy a "three-pronged test of legality, necessity and legitimate aims" to ensure restrictions

---

https://latamjournalismreview.org/articles/laws-and-bills-against-disinformation-in-latin-america/ for a summary of bills proposed across Latin America to stem the spread of disinformation online.

[35] *See* Daniela Kyle, *The Battle for Social Media Regulation: Can International Human Rights Bridge the Governance Gap in the Digital Space?*, UNIVERSAL RIGHTS GROUP NYC (June 28, 2021), https://www.universal-rights.org/universal-rights-group-nyc-2/the-battle-for-social-media-regulation-can-international-human-rights-bridge-the-governance-gap-in-the-digital-space/ (arguing international human rights law can help fill the "governance gap" regarding tackling social media disinformation).

[36] *See Status of Ratification Interactive Dashboard*, UNITED NATIONS HUMAN RIGHTS, https://indicators.ohchr.org (Select the International Covenant on Civil and Political Rights from the map filter to view State parties in Latin America).

[37] *See* International Covenant on Civil and Political Rights, art. 19(1–2), U.N. Doc. A/6316 (1966), 999 U.N.T.S. 171, *entered into force* 23 Mar. 1976 [hereinafter ICCPR], http://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx.

[38] *See* Irene Khan (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Disinformation and Freedom of Opinion and Expression*, para. 33, U.N. Doc. A/HRC/47/25 (Apr. 13, 2021), https://documents-dds-ny.un.org/doc/UNDOC/GEN/G21/085/64/PDF/G2108564.pdf?OpenElement.

[39] ICCPR art. 19(1–2).

"[do] not put in jeopardy the right itself."[40] "Legitimate aims" are comprised of a discreet list that includes protection of others' rights and reputations, "national security[,]… public order[,]… public health[, and]…morals."[41] Furthermore, any restrictions imposed for these reasons "may not put in jeopardy the right itself"[42] and "may never be invoked as a justification for the muzzling of any advocacy of multi-party democracy, democratic tenets and human rights."[43] As such, State parties assume both positive and negative duties with regard to the rights to freedom of opinion and expression, meaning they must both refrain from infringing on the right and take affirmative steps to give it effect.[44]

The rights to freedom of opinion and expression are additionally affirmed by instruments such as the Universal Declaration on Human Rights (UDHR). Article 19 of the UDHR guarantees "the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers."[45] The rights have also been recognized in the online context by a UNHRC Resolution that called on States to escalate efforts to protect the rights of freedom of opinion and expression in digital spaces.[46]

---

[40] *See* Khan, *supra* note 38, at para. 54; HUMAN RIGHTS COMMITTEE, *General Comment No. 10, Freedom of Expression (Art. 19)* (19th Sess. 1983), para. 1 (June 29, 1983) [hereinafter *CCPR General Comment 10*], https://www.ohchr.org/Documents/Issues/Opinion/CCPRGeneralCommentNo10.pdf (restrictions cannot endanger the right to freedom of expression as a whole).
[41] *See* Khan, *supra* note 38, at para 31 (freedom of opinion quote); *CCPR General Comment 10*, *supra* note 40, para. 1 (no limitations on freedom of opinion); ICCPR art. 19(3) (limitations).
[42] *See* HUMAN RIGHTS COMMITTEE, General Comment No. 34, *Article 19: Freedoms of Opinion and Expression* (102nd Sess. 2011), para. 21 (Nov. 29, 2011) [*hereinafter CCPR General Comment 34*], https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf.
[43] *See CCPR General Comment* 34 para. 23.
[44] *See* ICCPR, art. 2 (Positive obligations are assumed by the State Party to protect and give effect to the rights in the Covenant.).
[45] *See* G.A. Res. 217 (III) A, Universal Declaration of Human Rights art. 19 [hereinafter UDHR] (Dec. 10, 1948), https://www.ohchr.org/EN/UDHR/Pages/Language.aspx?LangID=eng.
[46] *See* Human Rights Council Res. 44/12, U.N. Doc. A/HRC/RES/44/12, at para. 2, 8(a) (July 16, 2020), https://documents-dds-ny.un.org/doc/UNDOC/GEN/G20/190/70/PDF/G2019070.pdf?OpenElement.

B. Freedom of Expression in the Inter-American System

Article 13 of the American Convention on Human Rights provides the following rights to Freedom of Thought and Expression, which augment the ICCPR framework with additional protections.

> 1. Everyone has the right to freedom of thought and expression. This right *includes freedom to seek, receive, and impart information and ideas of all kinds*, regardless of frontiers, either orally, in writing, in print, in the form of art, or *through any other medium of one's choice*.
>
> 2. *The exercise of the right provided for in the foregoing paragraph shall not be subject to prior censorship* but shall be subject to *subsequent imposition of liability*, which shall be expressly established by *law* to the extent *necessary* to ensure:
>
> a. *respect for the rights or reputations of others*; or
>
> *b.* the protection of *national security, public order, or public health or morals.*
>
> 3. The right of expression *may not be restricted by indirect methods* or means, *such as the abuse of government or private controls* over newsprint, radio broadcasting frequencies, or equipment used in the dissemination of information, *or by any other means* tending to *impede the communication and circulation of ideas and opinions*…[47]

[47] *See* Organization of American States, American Convention on Human Rights art. 13, Nov. 22, 1969, O.A.S.T.S. No. 36, 1144 U.N.T.S. 123 [hereinafter ACHR], https://www.cidh.oas.org/basicos/english/basic3.american%20convention.htm (emphasis added). To date, twenty-five countries have ratified the Convention. *See* OAS, *American Contention on Human Rights "Pact of San Jose, Costa Rica" (B-32)*, https://www.oas.org/dil/treaties_b-32_american_convention_on_human_rights_sign.htm.

Notably, Article 13 provides greater protection against censorship than the international instruments previously discussed due to its express prohibitions on prior censorship and indirect restrictions on the right.[48]

The Inter-American Commission further expounded on Article 13 through the *Declaration of Principles on Freedom of Expression*.[49] The Principles make clear that "[p]rior conditioning of expressions, such as truthfulness, timeliness or impartiality is incompatible with the right to freedom of expression recognized in international instruments."[50] As such, prior censorship, an intervention permitted in various European countries, is not compatible with the Inter-American legal standard.[51] The Convention's prohibition on indirect restrictions on freedom of expression also provides broad protection against state efforts to impose liability on intermediaries, like WhatsApp, "for third party posted content" or burdening processes vital to the press, such as "license renewals, nationality processes, [or] state publicity assignments," due to the chilling effect on expression that would result.[52]

Article 13 also imports the three-pronged test utilized by the ICCPR for assessing permissibility of restrictions on the right to freedom of expression into the regional legal framework.[53] The subsequent *Joint Declaration on Freedom of Expression and "Fake News", Disinformation, and Propaganda*, guidance issued by the United Nations and Organization of

---

[48] *See* Center for Studies on Freedom of Expression and Access to Information, *Content Moderation and Private Censorship: Standards Drawn from the Jurisprudence of the Inter-American Human Rights System* 3 (Dec. 2017) [hereinafter CELE], https://www.ohchr.org/Documents/Issues/Opinion/ContentRegulation/CELE.pdf.
[49] *See* Inter-Am. Comm'n H.R., Declaration of Principles on Freedom of Expression, (Oct. 2000), http://www.oas.org/en/iachr/expression/showarticle.asp?artID=26&lID=1
[50] *See id.* at para. 7; *see also* Special Rapporteurship for Freedom of Expression, *Background and Interpretation of the Declaration of Principles* para. 31, http://www.oas.org/en/iachr/expression/showarticle.asp?artID=132&lID=1 ("'erroneous,' 'untimely,' or 'incomplete'" information is protected).
[51] *See* CELE, *supra* note 48, at 2–3.
[52] *See* CELE, *supra* note 48, at 4.
[53] *See* ACHR art. 13(2).

American States alongside other regional bodies, reaffirmed that "the right… protects information and ideas that may shock, offend, and disturb."[54] The Declaration clarifies, however, that the three-pronged test implies that "[t]here should be no general or ambiguous laws on disinformation, such as prohibitions on spreading "falsehoods" or "non-objective information."[55] In other words, information that offends a government must not be prosecuted under a general disinformation law.

III. Nicaragua's Cybercrime Bill

This Part analyzes Nicaragua's Cybercrime Bill in light of the international and regional legal standards for restrictions on the right to freedom of expression and within the context of the 2018 popular protests against the Ortega-Murillo regime.

A. Movimiento 19 de Abril

In 2018, student-led protests across the country against social security reforms triggered a government crackdown on political opposition and galvanized calls for President Daniel Ortega's departure from office.[56] Documentation of protesters' deaths at the hands of police and paramilitary attacks ripped across the country on social media, decentralizing the call to protest.[57]

WhatsApp and Facebook represent the two most popular social networks in Nicaragua and have proved to be critical tools for disseminating news and protest updates.[58] The major telecommunications companies in Nicaragua, Claro and Movistar, offer zero-rating data deals that

---

[54] *See* Joint Declaration on Freedom of Expression and "Fake News", Disinformation and Propaganda, preamble, FOM.GAL/3/17, Mar. 3, 2017 [hereinafter Joint Declaration on Freedom of Expression], https://www.osce.org/files/f/documents/6/8/302796.pdf.
[55] *See id.* at 1(a)(iii)(3).
[56] *See* John L. Anderson, *"Fake News" and Unrest in Nicaragua*, NEW YORKER (Aug. 27, 2018), https://www.newyorker.com/magazine/2018/09/03/fake-news-and-unrest-in-nicaragua; Frances Robles, *In Just a Week, 'Nicaragua Changed' as Protesters Cracked a Leader's Grip*, N.Y. TIMES (Apr. 26, 2018), https://www.nytimes.com/2018/04/26/world/americas/nicaragua-uprising-protesters.html.
[57] *See* Courtney D. Morris, *Unexpected Uprising: The Crisis of Democracy in Nicaragua*, NACLA (May 14, 2018), https://nacla.org/news/2018/07/11/unexpected-uprising-crisis-democracy-nicaragua ("strategic use of social media" as a protest tool).
[58] *See* STATISTA, *Most Popular Social Networks Based on Share of Users in Nicaragua in 2018*, https://www.statista.com/statistics/754499/nicaragua-penetration-social-networks/.

typically include unlimited WhatsApp and Facebook usage. In late April 2018, news of the student protests, and documentation of the brutal police response, rippled out of Managua through WhatsApp, Twitter, and Facebook livestreams.[59] For example, video of a journalist, Ángel Gahona, being shot in the head while reporting on a Facebook livestream reverberated around the country, with many attributing his murder to the police.[60]

Political affiliation largely shaped the narrative people received on WhatsApp. While politically liberal people received news about violence against protesters, Sandinista WhatsApp groups labeled the protesters "terrorists," stoked by Vice President and First Lady Rosario Murillo's statements "denouncing the demonstrators as "tiny, petty, mediocre beings" and "vampires demanding blood."[61] Government supporters reworked the narrative to rally support behind Ortega to ensure "vandals" did not destroy the fruits of the Sandinista Revolution.[62]

Meanwhile, the Ortega-Murillo regime's efforts to rein in protests manifested in direct attacks on dissenting media, making WhatsApp critical for broadcasting information not aligned with the State narrative. Within days of the first protests, the government removed 100% Noticias, an independent news channel, from the air along with other independent media.[63] The government escalated their attacks on dissenting media throughout the summer of 2018, barring imports of ink and paper to stifle print publications by non-government-controlled media outlets, which arguably

---

[59] *See* John Otis, *In Nicaragua, Ortega's Control Over the Media Slips Even as a Government Crackdown Intensifies*, COMMITTEE TO PROTECT JOURNALISTS (Aug. 7, 2018), https://cpj.org/2018/08/in-nicaragua-ortegas-control-over-the-media-slips/.

[60] *See* Alexandra Ma, *A Journalist's Death was Captured on Facebook Live During Protests in Nicaragua*, BUSINESS INSIDER, Apr. 23, 2018), https://www.businessinsider.com/angel-gahona-killed-on-facebook-live-in-nicaragua-2018-4; *see also* Ángel Eduardo Gahona, COMMITTEE TO PROTECT JOURNALISTS, https://cpj.org/data/people/angel-eduardo-gahona/ (Although two young civilians were ultimately convicted for the murder, many distrust the trial's competence, as it was conducted in a courtroom closed to independent media.).

[61] *See* Anderson, *supra* note 56; ACHR art. 13(3). The Sandinista Party is the political party of President Daniel Ortega.

[62] *See* Richard E. Feinberg, *Nicaragua: Revolution and Restoration* 14, BROOKINGS (Nov. 2018), https://www.brookings.edu/wp-content/uploads/2018/11/FP_20181108_nicaragua.pdf.

[63] *See Nicaragua: Protests Leave Deadly Toll*, HUMAN RIGHTS WATCH (Apr. 27, 2018), https://www.hrw.org/news/2018/04/27/nicaragua-protests-leave-deadly-toll.

constitutes an illegal indirect restriction on expression under the American Convention on Human Rights standards.[64] As access to independent news dwindled, Nicaraguans increasingly relied on social media, including WhatsApp, to disseminate news and document human rights abuses by police and paramilitary forces.[65]

B. The Special Cyber Crimes Law

With social media having presented itself as a clear threat to Ortega's power, and with an election scheduled for November 7, 2021, Nicaragua's legislature passed the Special Cyber Crimes Law in October 2020.[66] The law prescribes two to four-year prison sentences for "those who promote or distribute false or misleading information that causes alarm, terror, or unease in the public," with offending information defined by the government.[67] If the information "incites hatred or violence, or puts at risk economic stability, public health, national sovereignty or law and order," defendants would instead be subject to three to five years of incarceration.[68] The broad phrasing of the law covers information distributed through traditional news outlets and social media.[69]

---

[64] *See* Kate Linthicum, *To Silence a Newspaper, Nicaragua's Government Took Away its Paper and Ink*, L.A. TIMES (Feb. 21, 2020), https://www.latimes.com/world-nation/story/2020-02-21/nicaragua-newspaper-president-ortega-la-prensa; CELE, *supra* note 48, at 4 (indirect restrictions).
[65] *See id.* (Articulo 66 disseminates news briefings on WhatsApp.).
[66] *See Nicaragua's Upcoming Election 'Has Lost All Credibility', US Says,* AL JAZEERA (Aug. 7, 2021), https://www.aljazeera.com/news/2021/8/7/nicaragua-upcoming-election-has-lost-all-credibility-us-says (Nicaragua's upcoming election); Associated Press, N*icaragua Approves "Cybercrimes" Law, Alarming Rights Groups*, PBS (Oct. 27, 2020), https://www.pbs.org/newshour/world/nicaragua-approves-cybercrimes-law-alarming-rights-groups.
[67] *See* Associated Press, *supra* note 65; Ley No. 1042, 2 Oct. 2020, Ley Especial de Ciberdelitos [Special Cybercrime Law] ch. IV, art. 30, LA GACETA, DIARIO OFICIAL [L.G.], 30 Oct. 2020 (Nicar.), http://legislacion.asamblea.gob.ni/normaweb.nsf/($All)/803E7C7FBCF44D7706258611007C6D87?OpenDocument.
[68] *See* Associated Press, *supra* note 65.
[69] *See* Ismael Lopez, *Nicaragua Passes Bill Criminalizing What Government Considers Fake News*, REUTERS (Oct. 27, 2020), https://www.reuters.com/article/nicaragua-politics/nicaragua-passes-bill-criminalizing-what-government-considers-fake-news-idUSL1N2HI2WG; Julieta Pelcastre, *Nicaragua to Punish Anti-Regime Opinions on Internet, Social Media*, DIÁLOGO (Nov. 30, 2020), https://dialogo-americas.com/articles/nicaragua-to-punish-anti-regime-opinions-on-internet-social-media/.

C. Assessment of Nicaragua's Compliance with its International and Regional Obligations Regarding Freedom of Expression

Nicaragua acceded to the International Covenant on Civil and Political Rights in 1980 without reservations.[70] Nicaragua has also ratified the American Convention on Human Rights on September 25, 1979, without reservations relevant to their Article 13 obligations.[71] As such, the State has international obligations under both treaties to refrain from infringing on freedom of expression, as measured by the three-part test of legality, necessity, and legitimate aims.[72] The Cybercrime Law likely fails this test on all fronts.

1. Legality

Legality requires that restrictions on the right to freedom of expression be enshrined in law and sufficiently clear as to the scope of prohibited conduct to "enable an individual to regulate his or her conduct accordingly," especially for laws invoking national security and public order justifications.[73] Under the Joint Declaration by the UN and OAS, "[g]eneral prohibitions on the dissemination of information based on vague and ambiguous ideas, including "false news" or "non-objective information," are incompatible with international standards for restrictions on freedom of expression… and should be abolished."[74]

Any infringement of rights to expression resulting from the Cybercrime Law are, ipso facto, enshrined in a law, however, the categories of offending information are broad and subject

[70] *See Status of Ratification Interactive Dashboard*, *supra* note 36, https://indicators.ohchr.org (Select Nicaragua from the drop-down menu to view the year of accession to the ICCPR.)
[71] *See* ACHR art. 13(2), *ratified by Nicaragua* September 25 1979, http://www.oas.org/dil/treaties_B-32_American_Convention_on_Human_Rights_sign.htm#Nicaragua (scroll down to view Nicaragua's declaration).
[72] *See ICCPR* art. 19(3); ACHR art. 13(2).
[73] *See General Comment 34* para. 24-25; Human Rights Council, Disease Pandemics and the Freedom of Opinion and Expression, para. 14 UN Doc. A/HRC/44/49 (Apr. 23, 2020), https://documents-dds-ny.un.org/doc/UNDOC/GEN/G20/097/82/PDF/G2009782.pdf?OpenElement (vagueness).
[74] *See* Joint Declaration on Freedom of Expression, *supra* note 54, at 2(a).

only to the government's interpretation, which is vulnerable to arbitrary implementation. As such, the law likely fails the legality prong due to vagueness.

2. Necessity

A restriction "violates the test of necessity if the… [legitimate purpose] could be achieved in other ways that do not restrict freedom of expression."[75] The Human Rights Council's *General Comment 27* clarifies that "restrictive measures must conform to the principle of proportionality; they must be appropriate to achieve their protective function; they must be the least intrusive instrument amongst those which might achieve their protective function; they must be proportionate to the interest to be protected…"[76]

Banning dissemination of all information the government deems false, including information shared on social media, is by no means the "least intrusive instrument" amenable to achieving the vague and numerous aims cited by the Special Cybercrime Law.[77] Additionally, the law is widely touted as targeting journalists for political opposition, and "penalization of a media outlet, publishers or journalist solely for being critical of the government or the political social system espoused by the government can never be considered to be a necessary restriction of freedom of expression."[78] Thus the law likely fails the test of necessity and proportionality.

3. Legitimate Aims

Under international law, there are two broad groupings of legitimate purposes for limiting freedom of expression, and no other aims are tolerable.[79] The first group is "respect for the rights or reputations of others" and the second is "protection of national security or of public order (ordre

[75] *See General Comment 34* para. 31.
[76] *See id.* at para. 34 (quoting General Comment 27 para. 14).
[77] *See id.*; Lopez, *supra* note 68.
[78] *See* 100% Noticias, *Nicaragua's "Gag Law" Takes Force*, HAVANA TIMES (Dec. 30, 2020), https://havanatimes.org/news/nicaraguas-gag-law-takes-force/; *General Comment 34* para. 42.
[79] *See General Comment 34* para. 22.

public), or of public health or morals."[80] "Suppress[ing]… from the public information of legitimate public interest" can never be a qualifying aim, and the Human Rights Committee calls for "extreme care" in the construction of treason laws to ensure compliance with Article 19(3) of the ICCPR and the overarching "aims and objectives of the Covenant."[81] The Committee further clarifies that limitations on the basis of morality cannot be subjectively defined and instead must be conceptualized within the international human rights framework.[82]

The purported purposes of the Cybercrimes law are to regulate false information and online sexual and financial crimes.[83] Facially, these aims may appear to fall within the approved categories of national security, public order, and morality. However, the broad construction of the law, its vulnerability to subjective interpretation by the government, and its convenience for suppressing information pertinent to the public interest could result in persecution of expression protected under Article 19 of the ICCPR in violation of the Covenant's intent. As such, the law likely fails this prong as well.

IV. Conclusion

Encrypted messaging applications like WhatsApp straddle the line between one-on-one conversation tools and a social network. Technical features like encryption and public group functionality, coupled with a user experience where social capital fosters trust, produces a unique atmosphere for disinformation to proliferate alongside true information dangerous to disseminate in more public forums.

Across Latin America, WhatsApp represents a key platform for digital expression because of its accessibility, affordability, and perceived privacy in repressive political climates like

[80] *See* ICCPR art. 19(3); ACHR art. 13(2).
[81] *See General Comment 34* para. 30, 26.
[82] *See id.* at para. 31.
[83] *See* Lopez, *supra* note 68.

Nicaragua, where signaling political dissent on more open platforms like Facebook is more easily detected. Because messaging applications have become critical to how individuals across Latin America communicate, efforts to curb disinformation on messaging applications must protect freedom of expression by observing the international and regional human rights standards of legality, necessity, and legitimate aims.

The Special Cybersecurity Law in Nicaragua fails each of the three prongs, constituting an illegal restriction on Nicaraguans' freedom of expression. The law exploits the growing global awareness about "disinformation" on social media to discredit political opposition and hamper the opposition social movement's efforts to disseminate information unfavorable to the Ortega-Murillo regime. The law's breadth, vagueness, and relinquishment of the definition of "false information" to the government facilitate convenient prosecutions of any expression with which the government does not agree. The Cybersecurity Law exemplifies a restriction on expression that "put[s] in jeopardy the right itself" by empowering prosecutors to not only use WhatsApp messages as evidence of crimes but as a crime in and of themselves and serves as an example of the self-serving abuse to which disinformation intervention laws are vulnerable if not strictly tied to international human rights standards.[84]

[84] *See CCPR General Comment 10* para. 4 (restrictions cannot endanger the right to freedom of expression as a whole). For information on how Ortega's government has used WhatsApp messages and WhatsApp group membership as evidence in criminal prosecutions of politically opposed individuals for charges of "conspiracy to undermine national integrity," see *Nicaragua: Trumped-Up Charges Against Critics*, HUMAN RIGHTS WATCH (Sept. 20, 2021), https://www.hrw.org/news/2021/09/20/nicaragua-trumped-charges-against-critics ("Trumped-up Charges" section) (WhatsApp messages used as evidence include messages sharing OAS resolutions on Nicaraguan elections and determining Nicaragua had not complied with OAS resolutions.).

Deepfake Technology in the United States and China: Disinformation and the Regulation of "Truth" in the Digital Age

I. Introduction

Deepfake technology has become a major regulatory concern for governments across the world. In the past few years, deepfake creations on the internet have generated heated debates around the impact of this new technology on information privacy and cybersecurity, as well as democratic principles such as free speech and political participation. A portmanteau of the terms "deep learning" and "fake,"[1] deepfake is shorthand for the range of hyper-realistic digital falsification of images, video, and audio.[2] Deepfake is an evolution of disinformation techniques that could manipulate information on a broadscale. Its ability to closely mimic real-world events makes it a powerful tool that could unsettle conceptions of "truth" in the digital age.

In both the United States and China, online platforms are under scrutiny to regulate the spread of deepfakes, although the sources of pressure differ.[3] In China, companies are responding to top-down legislations and directives from the national government that apply to all audiovisual service providers; whereas in the U.S., in the absence of federal rules, platforms are reacting to pressure from diverse stakeholders such as the media, civil society, and policymakers.[4] This chapter argues that the different approaches taken by China and the United States to regulate deepfakes raise distinct regulatory and free-speech concerns, which have significant implications

[1] Christopher Whyte, *Deepfake News: AI-Enabled Disinformation as a Multi-Level Public Policy Challenge*, J. OF CYBER POL'Y, 4 (2020).
[2] Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753, 1757 (2019).
[3] Julia Chen, *Deepfakes*, THE ASIA SOC'Y (Sep. 15, 2020), https://asiasociety.org/sites/default/files/inline-files/Final%20Deepfake%20PDF.pdf.
[4] *Id.*

on the construction and perception of "truth" in both societies. In the United States, First Amendment protections over false speech and political speech leave little room for the federal government to regulate non-consensual deepfakes and provide redress to harmed individuals.[5] In addition, Section 230 of the Communications Decency Act (CDA) provides wide immunity from liability to online platforms for hosting harmful content.[6] In China, public debates and regulatory responses have spurred important advances on practical issues such as fraud risks, image rights, economic profit, and ethical imbalances.[7] But the centralized and forceful approach of Chinese authorities to ban technologies and online speech that "aim at undermining the efficacy of authoritative discourse" further narrows online space where Chinese net-users could explore alternative versions of "truth" other than that sponsored by the Chinese Communist Party (CCP).[8]

The advent of deepfakes harkens an era when "truth" becomes more destabilized and contested than ever before. As Franklin Foer remarks, "we'll shortly live in a world where our eyes routinely deceive us … we're not so far from the collapse of reality."[9] But the "'truth' of audiovisual content has never been stable – truth is socially, politically, and culturally determined."[10] Governments and civil societies must contend with difficult technological, legal, and ethical questions as they regulate media manipulated by artificial intelligence (AI). Moving

[5] Shannon Reid, *The Deepfake Dilemma: Reconciling Privacy and First Amendment Protections*, 23 U. PA J. CONST. L. 209, 237 (2021).
[6] Chesney et al. , *supra* note 2, at 1795.
[7] De Seta, *supra* note 5, at 941.
[8] Qingning Wang, *Linguistic Violence and Online Political Communications in China: The Example of 鸡的屁 (Ji De Pi) as an Ironic Spoof of Gross Domestic Product in Online Debates Around Environmental Issues*, 3 GLOB. MEDIA AND CHINA 18, 22 (2018).
[9] Franklin Foer, *The Era of Fake Video Begins*, THE ATLANTIC (Sep. 26, 2021), https://www.theatlantic.com/magazine/archive/2018/05/realitys-end/556877/.
[10] Britt Paris & Joan Donovan, *Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence*, DATA & SOC'Y (Sep. 18, 2019), at 7, https://datasociety.net/wp-content/uploads/2019/09/DS_Deepfakes_Cheap_FakesFinal-1-1.pdf.

forward, we must question whether "those who claim or construct expertise around deepfakes will do so in a way that reinforces economic, political, and social privilege."[11]

The rest of the chapter will proceed as follow: after briefly discussing the basics of deep-fake technology in Part Two, Part Three will address regulatory responses from the United States government (or rather the lack thereof) and ways in which the government and platforms could better address the challenges that deepfake technology and the spread of disinformation pose on the democratic process. Part Four will then examine regulatory responses from the Chinese government and how they raise concerns about closing off online avenues for political expression. Lying at the heart of both cases is the inherent tension between mitigating harms caused by disinformation on the one hand and safeguarding free speech and political participation on the other.

II. Deepfakes as an Evolution in Disinformation Techniques

Deep-fake technology "leverages machine-learning algorithms to insert faces and voices into video and audio recordings of actual people and enables the creation of realistic impersonations out of digital whole cloth."[12] It involves the use of a "neural network" for machine learning. Just like how experience refines the brain's neural nodes, examples train the neural network system. By processing a broad array of training examples, the neural network can create increasingly accurate models. Through this learning process, neural networks categorize audio, video, or images to generate realistic impersonations or alterations.[13] Although large volumes of input data (*i.e.* legitimate imagery and video content) is required to make convincing deepfakes,

[11] *Id.* at 22.
[12] Chesney et al., *supra* note 2, at 1758.
[13] *Id.* at 1759.

that data is already widely available for public figures and is increasingly available for social media users.[14] As more photos of everyone circulate on the Internet, it will become relatively easy to create a deepfake video of anybody.[15]

While the use of new technology for broad-scoped information manipulation is nothing new, deepfakes endow falsehood with a "whole new, explosive emotional intensity."[16] Being applied to the digitization of bodies, including one's voice and likeness, deepfakes are most likely to have a negative impact on women, people of color, and those questioning powerful systems. The prevalence of deepfake-generated pornography is an example of how the new technology can be used to exploit a particular gender group.[17] As this new technology becomes more decentralized, people can spread manipulated information at new speeds and scales.[18] This increased distribution challenges traditional countermeasures to manipulation such as moderation or fact-checking.[19] Three phenomena – the information cascade dynamic, human attraction to negative and novel information, and filter bubbles – help to explain why deep fakes may be especially prone to going viral.[20] The information cascade dynamic describes how people stop paying sufficient attention to their own information and rely instead on what they assume others have reliably determined. In addition, people are more inclined to spread negative and novel falsehoods and information that confirm their pre-existing beliefs.[21] Thus, while deepfakes are an evolution of audiovisual medium manipulation, they present unique challenges for governments

[14] Whyte, *supra* note 1, at 10.
[15] *See supra* Introduction.
[16] Foer, *supra* note 17.
[17] Paris et al., *supra* note 12, at 7.
[18] *Id.*
[19] *Id.*
[20] Chesney et al., *supra* note 2, at 1765.
[21] *Id.*

and civil societies across the world trying to curb the spread of disinformation while keeping the Internet an open and free space for creative and democratic discourse.

III. Deep-fake Technology in the United States

In the United States, deepfakes have been used to create disinformation that destabilizes a shared conception of truth, which is the bedrock of political participation in a democratic society. However, any government intervention to regulate deepfakes would trigger First Amendment free speech concerns. Platform companies, which are probably in the best position to regulate deepfakes circulating online, are shielded from liability by the Communications Decency Act. How to check the spread of deepfake-generated disinformation without throttling free speech in a democratic society is a challenge for regulators moving forward.

Deepfakes have been put to pernicious use to distort media and reporting of politicians and social movements, putting the democratic process under pressure it has never experienced before. As Whyte remarks, democratic political systems are perhaps best thought of as information systems.[22] And democratic discourse is most functional when debate build upon a foundation of shared facts and truths supported by empirical evidence.[23] Deepfakes fundamentally undermine the stability of these shared facts and truths. In the infamous video of House of Representatives Speaker Nancy Pelosi, her speech had been slowed down by seventy-five percent, making it sound like she was slurring her words.[24] The edited version of the video, which was retweeted by the official Twitter account of President Trump, received over 6.3 million views. On Facebook,

[22] Whyte, *supra* note 1, at 10.
[23] Chesney et al., *supra* note 2, at 1765.
[24] *See* Lauren Feiner, *Facebook Says the Doctored Nancy Pelosi Video Used to Question her Mental State and Viewed Millions of Times Will Stay Up*, CNBC (May 24, 2019), https://www.cnbc.com/2019/05/24/fake-nancy-pelosi-video-remains-on-facebook-and-twitter.html.

commenters called Pelosi "drunk" and a "babbling mess."[25] In February 2018, in the aftermath of the horrific shooting at Marjory Stoneman Douglas High School in Parkland, Florida, one of the student survivors, Emma González, wrote an article for the *Teen Vogue* that included a collage of photos in which Emma rips up a large sheet displaying a bullseye target.[26] This powerful image that symbolized the emotional national debate on gun control was soon smeared by a falsified version, in which the torn sheet was changed from a bullseye to a copy of the Constitution of the United States.[27] Although the fakes were quickly flagged and criticized by journalists following the debate, they still distorted the national dialogue on gun control and casted a disturbing shadow on Parkland victims.

A. Free speech concerns limit the government's ability to regulate deepfakes

Despite the difficult and novel challenges that deepfakes pose on the democratic process, a general prohibition on deepfakes, even with guardrails, would pose serious concerns about limiting expression that is central to the American democratic culture.[28] In the landmark 1964 decision *New York Times v. Sullivan*, the Supreme Court held that false speech enjoys constitutional protection because its prohibition would deter truthful speech.[29] In 2012, in the plurality and concurring opinions of *United States v. Alvarez*, the Court went further to conclude that "falsity alone" does not remove expression from First Amendment protection.[30] Furthermore, laws forbidding political candidates' lies are especially prone to First Amendment challenges

[25] Stephen McDermott, *Explainer: Why is Facebook Allowing a Doctored Video of "Drunk" US Speaker Nancy Pelosi to Stay Online?, the journal* (May 26, 2019), https://www.thejournal.ie/nancy-pelosi-fake-news-video-facebook-4654225-May2019/.
[26] *Id.* at 1756.
[27] *Id.*
[28] *Id.* at 1789.
[29] N.Y. Times v. Sullivan, 376 U.S. 254 (1964).
[30] United States v. Alvarez, 567 U.S. 709, 736 (2012).

because political expression is vulnerable to government overreaching and partisan abuse.[31] The Court stated in *Brown v. Hartlage* that the "State's fear that voters might make an ill-advised choice does not provide the State with a compelling justification for limiting speech."[32] The First Amendment is thus a significant check on the government's ability to regulate deepfakes and provide harmed individuals with redress.[33]

B. Section 230 provides liability shield to platforms

If an outright ban is not legally feasible, could online platforms be held liable for the dissemination of deepfake-generated disinformation and falsehoods? The answer is probably no. In 1966, Congress provided platforms with a liability shield through Section 230 of the Communications Decency Act (CDA). Section 230(c)(1) forbids treating platforms as publishers or speakers of someone else's problematic content, thus shielding platforms from liability even if they encouraged the posting of that content.[34] Furthermore, to remove the disincentive to self-regulation, Section 230(c)(2) of the Act forbids civil suits against platforms based on the good-faith act of filtering to screen out offensive content, such as obscenity, harassment, or violence.[35] While Section 230 has enabled the Internet's early growth and benefited digital expression and democratic culture, it has also "evolved into a super-immunity that … prevents the best-positioned entities to respond to the most harmful content."[36] Under the CDA, platforms could ignore destructive activities, and could even solicit unlawful activities by ensuring that abusers cannot be

[31] Chesney et al., *supra* note 2, at 1803.
[32] 456 U.S. 45, 46 (1982).
[33] Reid, *supra* note 11, at 237.
[34] Chesney et al., *supra* note 2, at 1796.
[35] *Id.*
[36] *Id.* at 1798.

identified.[37] To a large extent, Section 230 ensures that platforms enjoy "power without responsibility."[38]

Given the constitutional restraint on the government, platforms are best-situated to regulate the dissemination of deep-fakes. Their terms-of-service (TOS) agreements "are the single most important documents governing digital speech in today's world."[39] Commentators have suggested modest adjustments to Section 230, either through judicial interpretation or legislation that would combine a reasonably calibrated standard of care with safeguards to reduce opportunities for abuses.[40] Platforms will continue to enjoy the protections of Section 230 so long as they take reasonable steps to ensure that their services are not being used to cause serious harm. These amendments will inevitably raise difficult questions regarding the standard of reasonableness, which would differ based on the nature and purpose of different online entities. The standard will also evolve as technology improves. But still, having these important policy conversations would be useful to incentivize stakeholders to mitigate harms flowing from user-posted or distributed deepfakes.

IV. Deepfake Technology in China

In contrast to the United States government, the Chinese government has taken a centralized and strict approach to regulate new technologies that could undermine the efficacy of authoritative discourse. The emergence of deepfakes in China could be viewed as an evolution of the longer history of vernacular creativity on the Chinese Internet since its creation. Practices of

[37] Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity*, 86 FORDHAM L. REV. 401, 423 (2017).
[38] Rebecca Tushnet, *Power Without Responsibility: Intermediaries and the First Amendment*, 76 GEO. WASH. L. REV. 986 (2008).
[39] Chesney et al., *supra* note 2, at 1817.
[40] *Id.* at 1800.

audiovisual manipulation such as image and video editing have given rise to repertoires of humor and parody shared online, commonly known as *egao*, or "mischievous spoofs."[41] Compared with users in a democratic setting, Chinese netizens must operate in a decentralized censorship scheme that combines coercive measures such as deleting posts and arresting dissidents with self-censorship.[42] Under such circumstances, creating online spoofs using novel technologies such as deepfakes has become a safer route of political and cultural expression.[43] Blanket government prohibition of deepfakes will inevitably chill these forms of alternative political discourse on the Chinese Internet. While privacy and disinformation are legitimate concerns, it remains to be seen whether Chinese netizens could continue to engage with deepfake technology in their creative discourse in a heavily censored online environment.

Deepfake captured nationwide attention in 2019 with the launch of the ZAO app, which allowed users to swap faces with celebrities and see themselves starring in their favorite movies or shows. This amusing app topped the topic ranks on Sina Weibo and became the most downloaded app on many app stores. Its servers crashed multiple times due to massive demand.[44] The app required users to scan their faces and agree to the app's terms of consent. However, within days, major newspapers published critical opinion pieces about ZAO's terms of consent, which seemed to grant its parent company Momo rights to store and reutilize user photos indefinitely or even sell them to third parties.[45] Under public pressure, ZAO had to revise its terms of consent,

[41] Gabriele de Seta, *Huanlian, or Changing Faces: Deepfakes on Chinese Digital Media Platforms*, 27 CONVERGENCE: INT'L J. OF RESCH INTO NEW MEDIA TECH. 935, 939 (2021).
[42] Bingchun Meng, *From Steamed Bun to Grass Mud Horse: E Gao as Alternative Political Discourse on the Chinese Internet*, 7 GLOB. MEDIA AND COMMC'N. 33, 39.
[43] *Id.*
[44] Gabriele de Seta, *Huanlian, or Changing Faces: Deepfakes on Chinese Digital Media Platforms*, 27 CONVERGENCE: INT'L J. OF RESCH INTO NEW MEDIA TECH. 935, 941 (2021).
[45] *Id.*

specifying that user photos would not be used for other purposes, and that users would be able to delete their personal information from the app's servers.[46]

A. Egao as political speech on the Chinese Internet

Because many critical political comments that directly discuss events or behaviors are prohibited, Chinese net-users find creative ways to play with discourse or mediums of expression on their own terms.[47] Probably the most famous example of a humorous play with discourse is the term "Grass Mud Horse" (*草泥马,* pronunciation is similar to fuck your mother). This vulgar and gendered term is also ironic and rhetorical. It was first used by Chinese net-users to criticize the government's censorship of online language and protest against the lack of freedom of expression.[48] Similarly, the term "Ji Di Pi" (*鸡的屁*, pronunciation is similar to chicken's fart and GDP) is used to criticize the government's GDP-driven policies that have been detrimental to the environment.[49] By using these creative terms that evade government censorship, Chinese netizens find ways to express their political opinions and communicate with each other in a scrutinized and controlled environment. Although deepfakes have not yet been widely used for political discourse, it belongs to this tradition of vernacular creativity on the Internet. Online content creators have already formed vibrant communities where they share, code, and comment on each other's work.[50] They also share content like textual walkthroughs and video tutorials aimed at introducing fellow

[46] *Id.*
[47] Wang, *supra* note 16, at 22.
[48] *Id.* at 19–20.
[49] *Id.*
[50] De Seta, *supra* note 46, at 948.

users to specific audiovisual synthesis tools, especially when they are hosted on platforms inaccessible from China like Google Colab.[51]

B. Crackdown on deepfake could limit avenues of political expression

The Chinese government has reacted quickly to crack down on creative speech that threatens to challenge the authoritative discourse. A directive from the State Administration for Press and Publications to "state departments for press, publication, radio, film and television" in 2018 stated that "governments at provincial, municipal and district level must engage in strict management of broadcasters, including mashups and remixes uploaded by internet users."[52] In the same year, the culture ministry also ordered the deletion of thousands of online videos for parodying popular "red classics and heroes."[53] Two weeks after the launch of the face-swapping app ZAO, the Cyberspace Administration of China (CAC) republished seven articles discussing face-swapping, including "a call for the development of more detailed AI regulation; a plea for the inclusion of 'humanistic and safety genes' in AI development …; several infographics illustrating the risks of [face-swapping]; and an overview of recent regulations on apps collecting personal data." On November 18, 2019, the State Internet Information Office (SIIO) announced the "Regulations on the Administration of Networked Audiovisual Information Services" to be enforced from January 1, 2020, which addresses "new technologies and new applications such as deep learning and virtual reality," directing platforms and providers to "ensure data security, flag synthetic content, and avoid publishing false information."[54] The CAC and the Public Security

[51] *Id.* at 947.
[52] Yang Fan, *China Bans Mashups, Spoofs and Re-Dubs of "Classic Literary Works" in Online Media*, RADIO FREE ASIA (Sep. 18, 2021, 1:06 PM), https://www.rfa.org/english/news/china/china-bans-mashups-spoofs-and-re-dubs-of-classic-literary-works-03232018120018.html.
[53] *Id.*
[54] Cyberspace Administration of China, Ministry of Culture and Tourism, & National Radio and Television Administration (2019) *关于印发《网络音视频信息服务管理规定》的通知 Guanyu yinfa "Wangluo yinshipin*

Bureau (PSB) met with eleven companies that use deep-fake technology in March 2021, including ByteDance, Alibaba Group Holding and Tencent Holdings, ordering these companies to conduct security reviews of their technology.[55]

While these regulations are commendable in protecting user data and preempting the spread of disinformation, they also threaten to stunt the development of a novel technology that could contribute to experimental and expressive discourse. By banning spoof or audiovisual manipulation that "distort" or "vilify" official rhetoric, the Chinese government is asserting its dominance as the shaper of discourse on the Chinese Internet, so that "truth" in that medium remains monolithic and uncontested.

V. Conclusion

This chapter has examined the emerging technology of deepfake and conducted a comparative analysis of the impact of deepfake in the United States and China. In the United States, disinformation spreading through deepfake can have a negative impact on the electoral process. However, constitutional protection over free speech and Section 230 of the CDC have chilled responses from both the government and platform companies to take adequate steps to mitigate harms caused by deepfake. In China, the CCP has taken proactive and aggressive steps to regulate this new technology, as well as other forms of digital expression that threatens to challenge the official discourse. Such aggressive measures threaten to throttle political speech through *egao*, a form of expression that has given Chinese netizens a channel of self-expression in a highly

xinxi fuwu guanli guiding" de tongzhi [Notice on the issuance of the "Regulations on the Administration of Internet Audiovisual Information Services"]. Cyberspace Administration of China. Available at: http://www.cac.gov.cn/2019-11/29/c_1576561820967678.html.

[55] Xinmei Shen, *Beijing Sharpens Focus on Deepfake Use, Social Audio Apps as It Pushes Security Review at Big Tech Companies*, SOUTH CHINA MORNING POST (Mar. 19, 2021), https://www.scmp.com/tech/policy/article/3126023/beijing-sharpens-focus-deepfake-use-social-audio-apps-it-pushes.

censored internet environment. More comparative studies in the future on how different governments attempt to contend with this technology will help to elucidate how "truth" is constructed and regulated by diverse stakeholders in disparate policy environments, and the social, legal, and ethical problems raised in such processe

The Prohibition of Internet Shutdowns in Africa by International Law

In 2020, the internet was intentionally disrupted more than 150 times across at least 29 different countries.[1] This trend has continued in 2021, with at least 50 shutdowns in 21 countries between January and May.[2] Though this number has decreased from the 115 shutdowns reported between January and May 2019, many of the more recent shutdowns have been longer.[3] For example, the internet has been cut off in the Tigray region of Ethiopia since November 2020.[4] Additionally, countries are using more sophisticated methods to block access to information via the internet, making shutdowns harder to detect.[5] For example, a government can create an internet blackout, by ordering internet service providers (ISPs) to block access completely.[6] However, authorities are increasingly targeting specific platforms, such as social media applications.[7] Alternatively, governments are using a technique called "throttling," which involves asking ISPs to slow down network traffic, sometimes to specific sites or applications.[8] Governments can also specifically target mobile internet connections.[9] Additionally, shutdowns can affect individual areas, such as a single village, or entire countries.[10] Of the shutdowns from January to May 2021, 24 affected a whole country or multiple states, provinces, or regions in a country; 11 affected more

[1] *See* Marianne Díaz Hernández, et al., *#KeepItOn Update: Who is Shutting Down the Internet in 2021?*, ACCESS NOW (June 7, 2021, 2:00 AM), https://www.accessnow.org/who-is-shutting-down-the-internet-in-2021/.
[2] *See id.*
[3] *See id.*
[4] *See id.*
[5] *See Internet Shutdowns Now 'Entrenched' in Certain Regions, Rights Council Hears*, UN NEWS (July 1, 2021), https://news.un.org/en/story/2021/07/1095142.
[6] *See* Joe Tidy & Becky Dale, *What Happens When the Internet Vanishes?*, BBC NEWS (Feb. 25, 2020), https://www.bbc.com/news/technology-51620158.
[7] *See Internet Shutdowns Have Become a Weapon of Repressive Regimes*, ECONOMIST (Oct. 15, 2021), https://www.economist.com/graphic-detail/2021/10/15/internet-shutdowns-have-become-a-weapon-of-repressive-regimes.
[8] *See id.*
[9] *See id.*
[10] *See id.*

than one city or area in the same state, province, or region; and 13 affected only one city, county, or village.[11] All of these internet shutdown techniques are more targeted than traditional internet blackouts, making them more difficult to detect. Further, because governments often ask telecommunications companies to shut down the internet or limit access, uncertainty over who actually ordered a shutdown may cause further difficulty.

Information shutdowns are most often justified by the need to fight fake news or hate speech.[12] Given the lack of accountability on social media sites, and difficulty in identifying sources of information on personal messaging applications, the need to control information may be a valid justification.[13] However, sources of legitimate information disappear during information blackouts, making it difficult for journalists and human rights advocates to discover and document events.[14] As a result, other sources flock to fill the information vacuum, often with disinformation. Therefore, information shutdowns create environments conducive to disinformation and are counterproductive tools for fighting disinformation.[15] In addition to fostering disinformation, internet shutdowns prevent people from expressing themselves freely, harm the economy, hinder students from attending lessons, prevent taxes from being paid, and cut people off from healthcare services.[16]

Ultimately, states have an obligation to leave internet and telecommunications services in place in most, if not all, instances due to several norms of international law that likely prohibit

[11] *See* Díaz Hernández et al., *supra* note 1. https://www.accessnow.org/who-is-shutting-down-the-internet-in-2021/
[12] *See* Yohannes Eneyew Ayalew, *Public International Law and Internet Shutdowns: Time to Unpack Emerging Norms?*, GRONINGEN J. INT'L L. (July 13, 2020), https://grojil.org/2020/07/13/public-international-law-and-internet-shutdowns-time-to-unpack-emerging-norms/.
[13] *See* Nishant Shah, *(Dis)information Blackouts: Politics and Practices of Internet Shutdowns*, 15 INT'L J. COMMC'N , 2693, 2694 (2021).
[14] *See* Felicia Anthonio et al., *Voices from Tigray: Ongoing Internet Shutdown Tearing Families, Communities, Businesses Apart*, ACCESS NOW (Sept.13, 2021), https://www.accessnow.org/voices-from-tigray-ongoing-internet-shutdown-tearing-families-communities-businesses-apart/.
[15] *See* Shah, *supra* note 13, at 2694.
[16] *See* Brian Stauffer, *Shutting Down the Internet to Shut Up Critics*, HUM. RTS. WATCH(2020), https://www.hrw.org/world-report/2020/country-chapters/global-5#.

information blackouts. This Chapter will discuss several of these norms and apply them to current internet shutdown cases. Part I addresses international telecommunications law, which generally grants states sovereign control over telecommunications in their jurisdiction, but creates several obligations that arise during shutdowns. Next, Part II discusses the international human rights regime, which prohibits shutdowns but leaves some room for states to restrict rights lawfully. Part III addresses how contract law can be applied to prohibit shutdowns when specific telecommunications companies are involved. Finally, Part IV applies these four norms two current cases of internet shutdowns in Ethiopia and Sudan.

I. International Telecommunications Law

International telecommunications law is strongly rooted in ideas of sovereignty, and therefore generally allows internet shutdowns. For example, the 1992 Constitution of the International Telecommunication Union lays out "the sovereign right of each State to regulate its telecommunication," in the preamble.[17] Additionally, Article 34(2) states that "[m]embers . . . reserve the right to cut off any . . . private telecommunications which may appear dangerous to the security of the State or contrary to its laws, to public order or to decency."[18] Article 35 also states that member states can "suspend the international telecommunication service . . . provided that it immediately notifies such action to each of the other Member States."[19] While internet shutdowns are likely permissible under international telecommunications law, states may be obligated to

[17] Constitution and Convention of the International Telecommunication Union, Dec. 22, 1992, S. Treaty Doc. No. 104–34, 1825 U.N.T.S. 330. All African countries have either ratified, accepted, approved, or acceded to the Constitution and Convention of the International Telecommunication Union.
[18] *Id.* art. 34(2).
[19] *Id.* art. 35.

justify these actions under utilitarian grounds such as protecting national security, public order, or public decency.[20] States are also obligated to notify other Member States.[21]

III. The International Human Rights Regime

The human rights law regime provides one basis for prohibiting internet shutdowns. The United Nations lays out several fundamental human rights that are relevant to internet shutdowns: the right to freedom of expression, the right to social security, the right to medical care, the right to education, and the right to work.[22] The right to freedom of expression includes the right to "seek, receive and impart information and ideas through any media and regardless of frontiers."[23] Therefore, the internet and other telecommunications services would be included as a media for expression. The United Nations later expressly recognized this when the United Nations Human Rights Council acknowledged that the internet is increasingly important to the exercise of these human rights, particularly of the right to freedom of expression.[24] Therefore, the Human Rights Council affirmed "that the same rights that people have offline must also be protected online."[25] The United Nations built upon this acknowledgement in another resolution in 2014, stating that "access to information . . . facilitates vast opportunities for affordable and inclusive education."[26] This idea was applied by the Indian High Court in *Shirin R.K. v. State of Kerala* where they held

[20] *See* Ayalew, *supra* note 12.
[21] *See* Ayalew, *supra* note 12.
[22] *See* G.A. Res. 217 (III) A, Universal Declaration of Human Rights (Dec. 10, 1948); International Covenant on Civil and Political Rights, *opened for signature* Dec. 16, 1966, S. Treaty Doc. No. 95–2, 999 U.N.T.S. 171 (entered into force March 23, 1976); International Covenant on Economic, Social and Cultural Rights, *opened for signature* Dec. 16, 1977, 993 U.N.T.S. 3 (entered into force Jan. 3, 1976). All but three African countries, South Sudan, Mozambique, and Botswana, are parties to the International Covenant on Economic, Social and Cultural Rights. The United Nations Declaration on Human Rights is not binding law in itself, but instead expresses fundamental values that are enshrined in other international instruments that are legally binding. It was adopted by consensus in the UN General Assembly and has been consistently invoked for more than sixty years, suggesting that it may be considered common law even absent another legally binding agreement.
[23] Universal Declaration of Human Rights, art. 19, Dec. 10, 1948, [treaty source]; International Covenant on Civil and Political Rights, art. 19, *opened for signature* Dec. 16, 1966, [treaty source] (entered into force March 23, 1976).
[24] *See* Human Rights Council Res., U.N. Doc. 20/8, at 1 (July 26, 2012).
[25] *Id.* at 2.
[26] Human Rights Council Res., U.N. Doc. 26/13, at 2 (July 14, 2014).

that restriction the use of mobile phones in a hostel infringes upon the right to education.[27] The Universal Declaration of Human Rights states that the exercise of human rights is "subject only to such limitations as are determined by law solely for the purpose of" recognizing the rights of others "and of meeting the just requirements of morality, public order and general welfare in a democratic society."[28] The International Covenant on Civil and Political Rights further states that "in time of public emergency which threatens the life of the nation and the existence of which is officially proclaimed," parties may "take measures derogating from their obligations . . . to the extent strictly required by the exigencies of the situation."[29] This means that any disruption of internet services must fulfil the strict requirements of legality, legitimacy, necessary, and proportionality.[30] Legality refers to the idea that an internet shutdown must be rooted in domestic law. Legitimacy means that an internet shutdown must be rooted in an enumerated human rights law. For example, the International Covenant on Civil and Political Rights states that the right to freedom of expression can only be restricted for the rights of others or for the protection of national security, public order, public health, or morals.[31] Additionally, the internet shutdown must be necessary to achieve that legitimate purpose. Finally, it must be proportional, in that the legitimate purpose must be of sufficient importance to justify the scope of the restriction of rights.[32]

In 2017, the ECOWAS Community Court applied the principle of legality and legitimacy to a case stemming from the Togolese government's actions cutting off access to the internet and

[27] *See* Shirin R.K. v. State of Kerala, W.P.(C.) 19716 of 2019, H.C. Ker., 19 Sept. 2019, 19, 24.
[28] G.A. Res. 217 (III) A, Universal Declaration of Human Rights, art. 29 (Dec. 10, 1948).
[29] International Covenant on Civil and Political Rights, art. 4, *opened for signature* Dec. 16, 1966, 993 U.N.T.S. 3 (entered into force Mar. 23, 1976). All African countries but South Sudan are Parties to the International Covenant on Civil and Political Rights.
[30] *See* Ayalew, *supra* note 12.
[31] *See* International Covenant on Civil and Political Rights, art. 19(3), *opened for signature* Dec. 16, 1966, 993 U.N.T.S. 3 (entered into force March 23, 1976).
[32] *See generally* Martin Luterán, *The Lost Meaning of Proportionality*, *in* PROPORTIONALITY AND THE RULE OF LAW: RIGHTS, JUSTIFICATION, REASONING 21–42 (Grant Huscroft, Bradley W. Miller, Grégoire Webber, eds., 2014).

telecommunications services during civil unrest. The court first found that internet access, while not a fundamental right itself, is a "derivative right" of the right to freedom of expression.[33] Therefore, the court stated that any interference with the right must be "provided for by the law specifying the grounds for such an interference."[34] The government justified their action based on national security, claiming that the protests had the "potential to degenerate into a civil war."[35] Though the court stated that national security could be a valid justification of derogating from the right to freedom of expression, there was no national legislation providing the means by which the right could be derogated from. Therefore, the shutdown, because it was not legal, was a violation of the right to freedom of expression.[36]

In *Womah Mukong v. Cameroon*, a case involving the arrest of a journalist and longtime opponent of the one-party system in Cameroon, the court stated that "the legitimate objective of safeguarding and indeed strengthening national unity under difficult political circumstances cannot be achieved by attempting to muzzle advocacy of multi-party democracy, democratic tenants and human rights."[37] Therefore, the court concluded that deciding which measures might meet the necessity test in these situations simply does not arise.[38] Under this argument, the majority of justifications used by States, including preventing disinformation, quelling protests, and even stopping cheating are all likely impermissible as unnecessary and disproportionate.[39] This is

[33] *See* Amnesty International Togo v. the Togolese Republic, Community Court of Justice of the Economic Community of West African States, ECW/CCJ/APP/61/18, Judgement, ¶ 11 (6 July 2020)
[34] *Id.*
[35] *Id.* ¶ 13.
[36] *See id.* ¶ 47.
[37] *Mukong v. Cameroon*, Communication No. 458/1991, UN Human Rights Committee (HRC) ¶ 9.7, 21 July 1994.
[38] *See id.*
[39] *See* SOUTHERN AFRICA LITIGATION CENTRE, NAVIGATING LITIGATION DURING INTERNET SHUTDOWNS IN SOUTHERN AFRICA 18 (2019).

especially true as the value of uninhibited expression is particularly high during times of public debate in a democratic society.[40]

Even a legal, legitimate, and necessary internet shutdowns would often be prohibited due to the lack of proportionality depending on the technique used. For example, a more targeted shutdown is more likely to be proportional, as it would restrict fewer rights and be more likely to meet a legitimate need. However, even targeted internet shutdowns hinder a number of different human rights such as freedom of expression, the right to medical care, the right to education, and more. Therefore, only the most targeted internet shutdowns are likely to be proportional. Further, total blackouts wipe out access to an almost innumerable number of individuals, and therefore may never be proportionate to a legitimate aim. In light of the lack of proportionality of any sweeping information blackout, The United Nations issued a joint resolution in 2015, stating that "filtering of content on the Internet, using communications 'kill switches' (i.e. shutting down entire parts of communication systems) and the physical takeover of broadcasting stations are measures which can never be justified under human rights law."[41] A subsequent resolution by the United Nations Human Rights Council condemned the use of deliberate internet shutdowns, even those that are more targeted, stating that "measures to intentionally prevent of disrupt access to or dissemination of information online," violates international human rights law."[42]

Even if an internet shutdown were to be legal, legitimate, and proportional, several regional instruments suggest that it would still be prohibited. The African Charter on Human and Peoples' Rights, which also contains provisions on the rights to free expression, work, health, and education,

[40] *See* International Covenant on Civil and Political Rights, *opened for signature* Dec. 16, 1966, S. Treaty Doc. No. 95–2, 999 U.N.T.S. 171 (entered into force March 23, 1976), General Comment 34, ¶ 34.
[41] JOINT DECLARATION ON FREEDOM OF EXPRESSION AND RESPONSES TO CONFLICT SITUATIONS ¶ 4(c), (May 4, 2015).
[42] Human Rights Council Res., U.N. Doc. 32/13, at 4 (July 1, 2016).

does not contain any derogation provisions, meaning that parties are not allowed to derogate their treaty obligations during public emergencies.[43] In a communication brought against Chad, the Commission stated that a civil war could not be used as a legal shield for failure to fulfill the legal obligations under the African Charter.[44] Additionally, the African Declaration of Principles on Freedom of Expression and Access to Information adopted by the African Commission on Human and Peoples' Rights in 2019 states unequivocally that "States shall not engage in or condone any disruption of access to the internet and other digital technologies for segments of the public or an entire population."[45] Therefore, internet shutdowns in African countries are unambiguously prohibited by international human rights law, regardless of the legality, legitimacy, or proportionality of the action.[46]

IV. Contract Law

In Sudan, Abdelazeem Hasaan took telecommunications company Zain to court after the internet was disrupted in the country.[47] Hasaan won the case under contract law, arguing that he entered into a service contract with Zain where he pays monthly fees for which the company is obliged to provide full services.[48] He asked the court for specific performance in terms of the contract and reserved his right to compensation.[49] Hasaan won the case, but because he filed in a

[43] *See* Commission Nationale des Droits de l'Homme et des Libertés v. Chad, Communication 74/92, African Commission on Human and Peoples' Rights [Afr. Comm'n H.P.R.] ¶ 40, (Oct. 1995)
[44] *See id.*
[45] Declaration of Principles on Freedom of Expression and Access to Information in Africa prin. 38(2), Nov, 10, 2019, African Commission on Human and Peoples' Rights [Afr. Comm'n H.P.R.], https://www.achpr.org/public/Document/file/English/Declaration%20of%20Principles%20on%20Freedom%20of%20Expression_ENG_2019.pdf.
[46] *See* Only one country, Burundi, has not signed or ratified the African Charter on Human and Peoples' Rights. Therefore, an information blackout in Burundi might be permitted so long as it was legal, legitimate, or proportional.
[47] *See Sudan Crisis: Internet Restored – but only for Lawyer*, BBC NEWS (June 24, 2019), https://www.bbc.com/news/world-africa-48744853.
[48] *See Kill Switch: Taking the Shutdown to Trial*, ACCESS NOW (Aug. 3, 2020), http://opentranscripts.org/transcript/kill-switch-taking-shutdown-to-trial/.
[49] *See id.*

personal capacity, Zain only restored access to his personal devices. Therefore, Hasaan returned to court with a class action and publicized his efforts.[50] He won again, and Zain uplifted the internet shutdown.[51]

Challenges rooted in contract law are particularly useful when an ISP will not acknowledge any government participation and when the government will not acknowledge any involvement. By bringing a suit against a service provider, the provider may disclose who authorized the shutdown. However, contract arguments are limited to the terms of the contract and on contract interpretation. For example, in a service contract from Botswana, the contract limits liability by stating "Orange Botswana shall not be liable to the Subscriber for any loss or damage suffered by the... subscriber whether same is direct or consequential, if . . . [t]he network Services are interrupted, suspended or terminated, for whatsoever reason."[52] Contracts like this may provide service providers the defense that they are not liable for service disruptions, and therefore stave off the need to raise a defense of reasonableness based on obeying a government order.[53]

VI. Ethiopia

Almost 40 years after a brutal fight in Tigray that contributed to massive starvation, the Ethiopian government is again launching a military offensive in Tigray. The government imposed an internet shutdown at the beginning of the conflict in the Tigray region in November 2020 including internet and telecommunication blackouts, and alleged blocking of specific websites.[54]

[50] *See id.*
[51] *See id.*
[52] TERMS AND CONDITIONS OF ORANGE BOTSWANA, Orange Postpaid Terms and Conditions ¶ 16.1, https://www.orange.co.bw/en/terms-conditions.html (last visited Nov. 5, 2021)
[53] *See* SOUTHERN AFRICA LITIGATION CENTRE, NAVIGATING LITIGATION DURING INTERNET SHUTDOWNS IN SOUTHERN AFRICA 37 (2019).
[54] *See* Ayalew, *supra* note 12.

As of September 12th, 2021, the Tigray region has been cut off from the rest of the world for over 300 consecutive days, with both broadband and mobile internet shut off.[55]

Despite the information blackout, information has slowly trickled from the region. The stories that often come months after the events in question took place have detailed numerous allegations of heinous crimes perpetrated by Ethiopian and Eritrean troops and the Amhara militia, including mass rape and sexual violence, mass murder, and abuse of refugees.[56] The information blackout has made it difficult for journalists and human rights advocates to discover and document human rights abuses, and has made it extremely difficult for people to connect with their families and maintain their livelihoods. Several twitter accounts, all originating in the United States, were created to fill the subsequent information vacuum with pre-written tweets that are then shared by other users. Much of the information in the pre-written tweets is unverified, and therefore potentially disinformation.

Ethiopia's internet shutdown is prohibited under all three norms of international law discussed here. First, though the Ethiopian government justified the internet shutdown citing civil unrest, they failed to notify other member states to the 1992 Constitution of the International Telecommunication Union. Further, though civil rest may make the internet shutdown legitimate, the shutdown is not legal because, just as in Togo, there is no national legislation providing the means by which the right to the freedom of expression could be derogated from. Additionally, because the internet shutdown has targeted an entire region consisting of over 7 million people and has gone on for over 300 days, there is not legitimate end that can make this sweeping blackout proportional and necessary. Finally, the African Declaration of Principles on Freedom of Expression and Access to Information, to which Ethiopia is a party, unambiguously prohibits

[55] *See* Anthonio, *supra* note 15.
[56] *See* Anthonio, *supra* note 15.

internet shutdowns. Telecommunication services are provided by the state-owned Ethio telecom, a monopoly that has frequently been responsible for both intentional and unintentional shutdowns.[57] In at least one previous instance the operator was sued for the outage, however the suit was based on human rights law.[58] Therefore, contract law is unlikely to provide a strong basis for prohibiting the shutdown in this case. Therefore, the strongest bases for seeking relief from the Ethiopian government for the current internet shutdown are international telecommunications law and by international human rights norms.

VII. Conclusion

Overall, there are several international legal norms that prohibit internet shutdowns. International telecommunications law is both outdated and rooted in sovereignty, however the Constitution and Convention of the International Telecommunication Union includes requirements to justify shutdowns and notify other Member States. The best defense against internet shutdowns is the international human rights regime, including international agreements such as the International Covenant on Economic, Social and Cultural Rights and the International Covenant on Civil and Political Rights, and regional instruments such as the African Charter on Human and Peoples' Rights. Though some of these agreements state that internet shutdowns are permissible when legal, legitimate, necessary, and proportional, no sweeping information blackout can be proportional due to the broad scope of the blackouts, and even more targeted internet shutdowns involve a number of human rights violations, including the right to freedom of expression, the right to health, the right to education, and more. Therefore, all but the most targeted internet shutdowns are disproportionate, and therefore prohibited under the international human rights regime. Finally,

[57] *See* Abdur Rahman Alfa Shaban, *Ethiopia Apologizes for Unexplained Internet Blackout, Customers Compensated*, AFRICA NEWS (June 19, 2019), https://www.africanews.com/2019/06/19/ethiopia-apologizes-for-unexplained-internet-blackout-customers-compensated//.
[58] *See id.*

contract law can be used to target specific service providers and to force a disclosure of government participation. This may be a prerequisite to bringing a suit against a State under international law.

Facebook and Accountability: How International Law is not Equipped to Hold Social Media Accountable

- A Case Study of Myanmar and the Rohingya -

I. Introduction

The mass displacement and atrocities inflicted on the Rohingya people in Myanmar has led to one of the biggest tragedies in recent history. In contrast to the atrocities of the 20th century, like the Holocaust and the Rwandan genocide, there is a new actor that has played an important role—Facebook. In the Rohingya conflict, Facebook as a social media platform has been weaponized by public figures and laypeople alike to spread hate speech and disinformation, which has fomented genocidal intent in the majority-Buddhist country of Myanmar.

In order to better understand the role Facebook played and the prospects for its accountability, this Chapter will first provide an overview of the historical context of the religious conflict between the ethnic Rakhine and the Rohingya. Then, it will evaluate the circumstances under which Facebook entered people's lives in Myanmar, and the effects the hate speech and disinformation have had on the Rohingya. Finally, this case study will evaluate the international proceedings that have been commenced against Myanmar in light of this crisis, namely the case before the International Court of Justice (ICJ) and the International Criminal Court's (ICC) investigation into the situation in Myanmar and Bangladesh, with an eye towards whether and how Facebook can be held accountable for the role it played in these atrocities.

II. Historical Context of the Persecution of the Rohingya

The presence of the Rohingya in the territory of Rakhine State in modern-day Myanmar dates back many centuries, even before British colonial rule. Historians have traced back the independent Rakhine kingdoms to antiquity, with the last of them established in 1430.[1] Mrauk U, the capital of the last Rakhine kingdom, thrived as a trade hub until its annexation by the Burmese in 1784-85. This annexation was not long lived in light of the First Anglo-Burmese War in 1824-26, which led to British control of the area.[2]

British colonialism facilitated the increase of the Muslim population in Rakhine, as Muslim workers from Bengal provided labor to expand rice cultivation in the area as part of colonial policies.[3] Some workers permanently settled in Rakhine, which changed the demographics of the region.[4] This change laid the foundation of religious strife between the Muslims and the ethnic Rakhines. Even though there have been sustained periods where the Muslim and Buddhist populations have lived in relative harmony, this harmony has been repeatedly disrupted by communal tension since the mid-19th century, especially after Myanmar's independence from colonial rule in 1948.[5]

The fact that Rakhine was left as a part of post-independence Myanmar disgruntled both Muslim and Buddhist populations in Rakhine and led to political resistance by both groups.[6] On the Muslim side, Rohingya Muslims led a rebellion in Rakhine, where they demanded equal rights

[1] *See* ADVISORY COMM'N ON RAKHINE STATE, TOWARDS A PEACEFUL, FAIR AND PROSPEROUS FUTURE FOR THE PEOPLE OF RAKHINE: FINAL REPORT OF THE ADVISORY COMMISSION ON RAKHINE STATE 18 (2017), https://www.rakhinecommission.org/app/uploads/2017/08/FinalReport_Eng.pdf.
[2] *Id*.
[3] Mohammad Shahabuddin, *Post-colonial Boundaries, International Law, and the Making of the Rohingya Crisis in Myanmar*, 9 ASIAN J. INT'L L. 334, 354 (2019).
[4] ADVISORY COMM'N ON RAKHINE STATE, *supra* note 1, at 18.
[5] *Id*. at 19.
[6] *Id*.

and an autonomous Muslim region in the north of Rakhine.[7] During the Second World War, they also aligned themselves with the British, whereas the Rakhine Buddhists were on the Japanese side.[8] Burmese nationalists also claimed that the Rohingya formed their own army around the time of independence and also approached Muhammad Ali Jinnah, Pakistan's first governor-general, to incorporate Northern Rakhine into East Pakistan (present-day Bangladesh).[9] The Burmese ruling elite used these allegations to question the Rohingyas' political allegiance to Myanmar, denying them citizenship, and referring to them as "Bengali foreigners."[10] Thus, the foundation of the widespread atrocities of 2017 was laid long before.

III. Proliferation of Facebook in Myanmar

Myanmar transitioned to civilian rule after decades of being ruled by the military in 2011.[11] The military rule meant that freedom of expression and speech were suppressed to such an extent that the United States and other European countries imposed sanctions, which led to the country's isolation from the rest of the world.[12] The arrival of civilian rule brought with it, *inter alia*, relaxed media censorship, which meant that the public was also able to more readily access telecommunications services, including purchasing a SIM card and a smart phone.[13] Where SIM cards used to cost around $200, the general public in Myanmar was now able to purchase a SIM

[7] Shahabuddin, *supra* note 3, at 357.
[8] Hannah Beech, *Across Myanmar, Denial of Ethnic Cleansing and Loathing of Rohingya*, N.Y. TIMES (Oct. 24, 2017), https://www.nytimes.com/2017/10/24/world/asia/myanmar-rohingya-ethnic-cleansing.html.
[9] MARTIN J. SMITH, BURMA: INSURGENCY AND THE POLITICS OF ETHNICITY 41 (1991).
[10] Shahabuddin, *supra* note 3, at 357.
[11] Kenneth Roth, *World Report 2012: Burma*, HUMAN RIGHTS WATCH (2012), https://www.hrw.org/world-report/2012/country-chapters/burma.
[12] Sheera Frenkel, *This Is What Happens When Millions Of People Suddenly Get The Internet*, BUZZFEED NEWS, (Nov. 20, 2016), https://www.buzzfeednews.com/article/sheerafrenkel/fake-news-spreads-trump-around-the-world.
[13] NEHGINPAO KIPGEN, MYANMAR: A POLITICAL HISTORY 78 (2016).

card for the equivalent of $2.[14] This sharp decline of prices led to a surge of smartphone usage and the exponential proliferation of social media, particularly Facebook.[15]

Due to the simultaneous introduction of the internet and Facebook into people's lives and poor internet literacy, "internet" and "Facebook" became synonymous to the public.[16] This created the perfect environment for the spread of disinformation and hate speech which stirred up underlying religious conflict, without any fact-checking. The Rohingya were at the receiving end of this online discourse, with millions finally having a platform to voice their anti-Rohingya sentiment without any inhibitions. The platform was also harnessed by community leaders who led the charge with hate speech, one of them being a Buddhist monk named Ashin Wirathu—called the "Face of Buddhist Terror" by Time magazine—who used his Facebook page to stoke fears against the Rohingya.[17]

Facebook's response to the spread of disinformation and hate speech directed at the Rohingya has been lethargic, at best. According to a 2018 Reuters research, there were more than 1000 examples of posts, comments, and pornographic images targeting the Rohingya on Facebook, some of which have been up on the platform for as long as six years.[18] A big reason for this delay has been the fact that Facebook had not recruited enough Burmese speaking content moderators to track and take down such posts on time; it only had four Burmese speaking content moderators in 2015, compared to 7.3 million Burmese-speaking Facebook users.[19] Another reason has been

[14] Anisa Subedar, *The country where Facebook posts whipped up hate*, BBC (Sept. 12, 2018), https://www.bbc.com/news/blogs-trending-45449938.
[15] Brad Ridout et al., *Social Media Use by Young People Living in Conflict-Affected Regions of Myanmar*, 23 CYBERPSYCHOLOGY, BEHAV., & SOC. NETWORKING (2020).
[16] *Id*.
[17] Subedar, *supra* note 14. *See also* Marella Oppenheim, *'It only takes one terrorist': the Buddhist monk who reviles Myanmar's Muslims*, THE GUARDIAN (May 12, 2017), https://www.theguardian.com/global-development/2017/may/12/only-takes-one-terrorist-buddhist-monk-reviles-myanmar-muslims-rohingya-refugees-ashin-wirathu.
[18] Steve Stecklow, *Why Facebook is losing the war on hate speech in Myanmar*, REUTERS (Aug. 15, 2018), https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/.
[19] *Id*.

the fact that Facebook's system for reporting posts was also in English, a language that is not widely spoken in Myanmar.[20]

IV. The Effects of Hate Speech and Disinformation

The proliferation of hate speech and disinformation on Facebook did not come without consequences. In July 2014, riots that lasted four days broke out in the city of Mandalay after unverified rumors of the alleged raping of a Buddhist employee by her Muslim employers spread online.[21] The riots resulted in the death of two men: a Buddhist and a Muslim.[22]

On a much larger scale, this proliferation entrenched the public's apathy and even malice to the Rohingya's plight. In August 2017, in the aftermath of small-scale attacks from the militants of Arakan Rohingya Salvation Army, the Burmese military, known as the Tatmadaw, conducted brutal and disproportionate counterinsurgency campaigns in Rohingya villages in Rakhine, leading to the mass killings and displacement of civilian Rohingyas, which the UN High Commissioner for Human Rights called a "textbook example of ethnic cleansing."[23] Despite concrete evidence of the atrocities that took place in Rohingya villages, including satellite images of more than 200 burned-down villages,[24] consistent accounts of extrajudicial killings and rape of civilians,[25] and

---

[20] *Id*.
[21] Mong Palatino, *The Meaning of the Mandalay Riots in Myanmar*, THE DIPLOMAT (July 12, 2014), https://thediplomat.com/2014/07/the-meaning-of-the-mandalay-riots-in-myanmar/.
[22] *Id.*
[23] UNITED NATIONS OFFICE OF THE HIGH COMMISSIONER OF HUMAN RIGHTS, DARKER AND MORE DANGEROUS: HIGH COMMISSIONER UPDATES THE HUMAN RIGHTS COUNCIL ON HUMAN RIGHTS ISSUES IN 40 COUNTRIES (2017) https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=22041&LangID=E.
[24] Sergio Peçenha & Jeremy White, *Satellite Images Show More Than 200 Rohingya Villages Burned in Myanmar*, N.Y. TIMES (Sept. 18, 2017), https://www.nytimes.com/interactive/2017/09/18/world/asia/rohingya-villages.html?_r=0.
[25] Hannah Beech, *Desperate Rohingya Flee Myanmar on Trail of Suffering: 'It Is All Gone'*, N.Y. TIMES (Sept. 2, 2017), https://www.nytimes.com/2017/09/02/world/asia/rohingya-myanmar-bangladesh-refugees-massacre.html?action=click&contentCollection=Asia%20Pacific&module=RelatedCoverage&region=Marginalia&pgtype=article&_r=0.

more than 900,000 Rohingya refugees currently seeking asylum in neighboring countries,[26] public opinion in Myanmar was not swayed.

The common narrative among politicians, religious leaders, and even local human rights activists in Myanmar has been that the Rohingya are not rightful citizens and are falsely trying to "hijack the world's sympathy" through the "power of a resurgent Islam."[27] All these different groups have responded to the United Nation's accounts by saying that the Rohingya were doing the mass burnings and targeting of civilians to themselves.[28] Myanmar's *de facto* leader at the time Aung San Suu Kyi went so far as to deny the Rohingya genocide when she took the stand in the ICJ case brought by The Gambia, which led to an outpour of support for her back in Myanmar.[29] These messages have inevitably found their way to Facebook and reached millions of people that perpetuate this denialism. They have also led to a distrust and skepticism of international aid groups and media, who the public viewed as "on the side of the terrorists."[30]

The current situation that the Rohingya are facing is further complicated by the military coup that took place in February 2021, which placed the Tatmadaw—the main perpetrators of the genocide—in power.[31] Even though there were talks of repatriation before the coup and the current coup leaders have vowed to "protect" the Rohingya as repatriation goes forward, the Rohingya are now even more apprehensive about returning to a Myanmar that is under a military leadership.[32]

---

[26] USA FOR UNHCR, ROHINGYA REFUGEE CRISIS EXPLAINED (2021), https://www.unrefugees.org/news/rohingya-refugee-crisis-explained/.
[27] BEECH, *supra* note 8.
[28] *Id*.
[29] Maung Zarni, *Why Myanmar's genocide denial will come back to haunt it*, WASH. POST (Jan. 15, 2020), https://www.washingtonpost.com/opinions/2020/01/15/aung-san-suu-kyi-must-be-held-account/.
[30] *Id*.
[31] Alice Cuddy, *Myanmar coup: What is happening and Why?*, BBC (Apr. 1, 2021), https://www.bbc.com/news/world-asia-55902070.
[32] Ashley Westerman, *What Myanmar's Coup Means For The Rohingya*, NPR (Feb. 11, 2021), https://www.npr.org/2021/02/11/966923582/what-myanmars-coup-means-for-the-rohingya.

The public opinion on the Rohingya has not shifted in the interim, so it remains to be seen whether repatriation efforts can take place safely in the current political climate.

V. Accountability Gap: How International Law is not Equipped to hold Facebook Accountable

The occurrence of another genocide before the eyes of the world led to conversations about how to hold those accountable for the crimes committed. However, unlike historical examples of genocide, Facebook comes in as a unique player that international law may not be equipped to address yet. Unlike a traditional news source, Facebook does not create its own news content and merely facilitates the use of its platform by ordinary citizens.

This section will provide an overview of the current international proceedings, namely the case against Myanmar before the ICJ and the ongoing investigation of Myanmar by the ICC, to determine whether and how Facebook's actions feature in these proceedings. The section will demonstrate how the current state of public international law and international criminal law is ill-equipped to bring accountability to Facebook.

A. Proceedings before the ICJ: The Gambia vs. Myanmar

In 2019, The Gambia began proceedings against Myanmar before the ICJ, alleging that the government of Myanmar's actions against the Rohingya constituted violations of the Genocide Convention.[33] In its submission, The Gambia asked the Court, *inter alia*, to institute provisional measures aimed at protecting the rights of the Rohingya under the Genocide Convention and to prevent Myanmar from aggravating or prolonging the dispute pending the ICJ's final judgment.[34]

[33] Press Release, INT'L CT. JUST., The Republic of The Gambia institutes proceedings against the Republic of the Union of Myanmar and asks the Court to indicate provisional measures (Nov. 11, 2019), https://www.icj-cij.org/public/files/case-related/178/178-20191111-PRE-01-00-EN.pdf
[34] *Id*. at 2.

To address *prima facie* jurisdiction, standing, and provisional measures, the ICJ issued an order on January 23, 2020.[35] With respect to jurisdiction, it found that there was a dispute within the meaning of Article IX of the Convention that allowed The Gambia, as a state party to the Genocide Convention, to bring a claim against Myanmar, who is also a state party to the Convention.[36] The ICJ also found that The Gambia had standing to bring the case, reasoning that the obligations that arise from the Genocide Convention are *erga omnes partes* and that a state party need not be specially affected by the actions of another state party in order to bring a claim under the Genocide Convention.[37]

The ICJ also found that The Gambia's allegations demonstrates that there is a real and imminent risk of irreparable prejudice to the rights invoked by The Gambia, which meant that the ICJ could put provisional measures in place.[38] Thus, the ICJ ruled that Myanmar must take measures to prevent the commission of all acts within the scope of Article II of the Genocide Convention.[39] It also ruled that Myanmar must take measures to prevent the destruction and ensure the preservation of any evidence regarding the allegations of acts within the scope of Article II and must report on all the measures taken pursuant to the order.[40] The case is currently pending further submissions by the parties.

Although Facebook has featured in The Gambia's pleading, especially in the context of its allegations regarding hate speech and disinformation,[41] the ICJ is not a forum where there can be

[35] Application of the Convention on the Prevention and Punishment of the Crime of Genocide (Gam. V. Myan.), Provisional Measures, 2020 I.C.J 3 (Jan. 23).
[36] *Id*. at ¶ 31.
[37] *Id*. at ¶¶ 41–42.
[38] *Id*. at ¶¶ 74–75.
[39] *Id*. at ¶ 79.
[40] *Id*. at ¶¶ 81–82.
[41] Application of the Convention on the Prevention and Punishment of the Crime of Genocide (Gam. V. Myan.), Application Instituting Proceedings and Request for Provisional Measures, ¶ 45 (Nov. 11, 2019), https://www.icj-cij.org/public/files/case-related/178/178-20191111-APP-01-00-EN.pdf.

any relief sought against Facebook itself. ICJ is only a forum for state-to-state dispute resolution, and it cannot entertain claims against private actors.[42] Thus, the only way Facebook can feature in further proceedings before the ICJ is by providing evidence of the hate speech and disinformation that is alleged to have taken place on the platform, and there has been an effort by The Gambia to facilitate the release of such information. Most recently, pursuant to The Gambia's request, a judge in the US District Court for the District of Columbia ordered Facebook to release records of accounts linked to violence against the Rohingya.[43] In his opinion, the judge stated that "[l]ocking away the requested content would be throwing away the opportunity to understand how disinformation begat genocide of the Rohingya and would foreclose a reckoning at the ICJ."[44] Although this is a laudable step in further exposing how Facebook was utilized to incite a genocide, it does little with respect to holding Facebook accountable in a state-to-state dispute.

B. ICC's Investigation of Myanmar and Bangladesh

In 2018, pursuant to the ICC Prosecutor's request for a decision on jurisdiction, the ICC found that it may assert jurisdiction pursuant to Article 12(2)(a) of the Rome Statute if at least one element of a crime within ICC's jurisdiction or part of such crime is committed on the territory of a state party.[45] Thus, the deportation of the Rohingya to Bangladesh, a party to the Rome Statute, meant that an element of the crime in question—deportation as a crime against humanity—occurred in a state party. This decision came under the vehement opposition of Myanmar, which stated that it could not be subject to ICC's jurisdiction as a non-party.[46]

---

[42] Article 34(1) of the ICJ Statute: "Only states may be parties in cases before the Court."
[43] Poppy Mcpherson, *U.S. court orders Facebook to release anti-Rohingya content records for genocide case*, REUTERS (Sept. 23, 2021), https://www.reuters.com/business/media-telecom/us-court-compels-facebook-release-records-anti-rohingya-content-report-2021-09-23/.
[44] *Id*.
[45] Decision on the "Prosecution's Request for a Ruling on Jurisdiction under Article 19(3) of the Statute", ICC-RoC46(3)-01/18-37 06-09-2018 1/50 RH PT ¶¶ 78–79 (Sept. 6, 2018) [hereinafter Jurisdiction Decision].
[46] *Id*. at ¶ 35.

In ruling on jurisdiction, the ICC first found that it had an objective legal personality, which brings with it the capacity to act (complementary to national jurisdictions) against impunity for the most serious crimes that concern the international community as a whole.[47] However, the ICC was careful not to say that this meant that it had automatic or unconditional *erga omnes* jurisdiction.[48] It evaluated the extent of its jurisdiction with respect to deportation as a crime against humanity in order to determine if it had jurisdiction over Myanmar under Article 12(2)(a).[49] It concluded that the object and purpose of Article 12(2)(a) supports the finding that the ICC can exercise jurisdiction when one element of the crime of deportation or part of it is committed on the territory of a state party.[50] In particular, it emphasized the "inherently transboundary nature" of the crime, that the conduct necessarily takes place on the territories of at least two states, and that the Rome Statute did not limit the applicability of the provision to deportations from one state party to another state party.[51] Thus, it concluded that acts of deportation that were initiated in a non-state party and completed in a state party fell within ICC's jurisdiction.[52]

Once the ICC made this jurisdictional ruling, it authorized the initiation of an investigation into the situation in Myanmar and Bangladesh in 2019.[53] Currently, the investigation is ongoing and it remains to be seen whether and against whom the Prosecutor will bring charges. Unlike the ICJ's mandate, the purpose of international criminal law and the ICC is to ensure individual criminal responsibility for committing crimes within its jurisdiction, which means that private individuals can be prosecuted. However, this may not be enough to hold Facebook accountable as

[47] *Id.* at ¶ 48.
[48] *Id.* at ¶ 49.
[49] *Id.* at ¶ 50.
[50] *Id.* at ¶ 70.
[51] *Id.* at ¶ 71.
[52] *Id.* at ¶ 73.
[53] Decision Pursuant to Article 15 of the Rome Statute on the Authorisation of an Investigation into the Situation in the People's Republic of Bangladesh/Republic of the Union of Myanmar, ICC-01/19 (Nov. 14, 2019), https://www.icc-cpi.int/CourtRecords/CR2019_06955.PDF [hereinafter Authorization Decision].

a private actor. Article 25(1) of the Rome Statute clearly states that the ICC has jurisdiction over "natural persons" and does not mention corporations.[54] Even if that hurdle was somehow surpassed, Facebook's slow response to taking down hate speech and disinformation does not fit within any of the actions under Article 25(3), since it did not create such content, nor can it be argued that its actions were done to further criminal activity.

In this regard, it is instructive to take stock of the experience of other international criminal tribunals that have dealt with somewhat analogous situations where media companies were featured in incitement of genocide charges. One such case is the Nahimana case (also known as the Media case) before the International Criminal Tribunal for Rwanda (ICTR), which involved three defendants who were purportedly the "masterminds behind a media campaign to desensitize the Hutu population and incite them to murder the Tutsi population in Rwanda in 1994."[55] Two of the defendants founded a radio channel that broadcasted anti-Tutsi messages nationwide from July 1993 to July 1994.[56] The remaining defendant managed a newspaper that published similar content consisting of "hate-filled messages" about the Tutsis between 1990 and 1995.[57] The Trial Chamber convicted all three for incitement and the Appeals Chamber affirmed.[58]

An important distinction between the Media case and the situation of Facebook is that all the defendants in the Media case were in a managerial position with direct control over the content that was being published. Thus, the content that was published and broadcasted could be readily attributable to the defendants in question. This is not the case with Facebook, which does not itself hold such an editorial function, other than taking down content that contravenes its own company

[54] *Id.*
[55] Sophia Kagan, *The "Media Case" Before the Rwanda Tribunal: The Nahimana et al. Appeal Judgement*, THE HAGUE JUST. PORTAL (2008).
[56] *Id.*
[57] *Id.*
[58] Neema Hakim, *How Social Media Companies Could Be Complicit in Incitement to Genocide*, 21 CHI. J. INT'L L. 83, 98 (2020).

policies. This *post hoc* review is different than initially giving its approval for certain content to be posted. As a result, even if public figures like Ashin Wirathu could conceivably be prosecuted for incitement of genocide by using Facebook as a medium to get their messages out, the current state of international criminal law makes it unlikely that the management of Facebook itself would be held accountable for incitement directly.

VI. Conclusion

The ongoing international proceedings before the ICJ and ICC provides some hope for accountability for those that are most directly responsible for the atrocities inflicted on the Rohingya. However, as this Chapter demonstrates, neither of the current international law paradigms for accountability provides satisfactory options for holding Facebook accountable for creating the infrastructure in which hate speech and disinformation spread exponentially. Nevertheless, this should not be cause for despair—necessity has been the driving force behind both international and domestic rulemaking. Given that this is only one case study among many against Facebook, and there is now more political will to hold the company accountable, states are now more incentivized to create laws that will lead to accountability. In the international sphere, this can manifest itself as maybe revisiting the Rome Statute to include criminal accountability of corporations and online platforms or even negotiating new treaties that specifically target financial and/or criminal liability for such corporations or platforms. Whatever the methodology employed, it is important that this issue is expediently addressed in light of the rapid developments in technology that allowed Facebook to wield such power in the first place.

Disinformation in an Information Vacuum: How the Criminalization of the Free Press Allows Countries to Create More Effective Disinformation Campaigns

Introduction

To support disinformation campaigns, states often work to create an informational vacuum through the criminalization of journalism. This elimination of trustworthy news sources from a region creates an environment ripe for social media conspiracy theories and state-sponsored disinformation campaigns. Thus, through the criminalization of journalism, states are able to conceal the truth about internal country conditions to the outside world and conceal the truth about the outside world to the state's citizens. This allows their narratives to gain greater traction by eliminating competition and fact-checking from outside news sources, allowing disinformation campaigns conducted through social media and state propaganda to achieve greater success. These techniques have been used in a wide range of countries. This chapter will discuss their usage in Myanmar during the Rohingya crisis and the 2021 coup d'état, Ethiopia's Tigray conflict, and North Korea.

I. Myanmar

A. Rohingya Crisis

Throughout the Rohingya crisis in Myanmar's Northern Rakhine State, the Burmese government worked to prevent news about the crisis from leaving the country's borders. To accomplish this goal, the country jailed journalists reporting on the situation. In late 2017, Reuters journalists Wa Lone and Kyaw Soe Oo were working on an investigative report on the massacre

of several Rohingya in the Northern Rakhine State's Inn Din village.[1] Before these journalists were able to release their report, the Burmese police arrested the journalists, who were subsequently charged and convicted of possession of secret national security documents under the Official Secrets Act.[2]

The Official Secrets Act was passed when Myanmar was a British colony and was broadly intended to criminalize the sharing of any information held by the colonial government with prison sentences of up to fourteen years.[3] The language of this statute is extremely broad, criminalizing even "approach[ing]…any prohibited place"[4] while defining "any prohibited place" to encompass a wide range of locations with even tenuous links to national security.[5] Furthermore, under § 3(2), the statute allows the government to prosecute individuals so long as it "appears that his purpose was a purpose prejudicial to the safety or interests of the State."[6] To prove purpose, the statute also permits the government to consider a defendant's "known character," which has allowed the law to be used to target individuals who are critical of the government.[7]

---

[1] Wa Lone, Kyaw Soe Oo, Simon Lewis & Antoni Slodkowski, *How Myanmar forces burned, looted and killed in a remote village*, REUTERS (Feb. 8, 2018), https://www.reuters.com/investigates/special-report/myanmar-rakhine-events/; Simon Lewis & Shoon Naing, *Two Reuters reporters freed in Myanmar after more than 500 days in jail*, REUTERS (May 6, 2019), https://www.reuters.com/article/us-myanmar-journalists/two-reuters-reporters-freed-in-myanmar-after-more-than-500-days-in-jail-idUSKCN1SD056.

[2] Simon Lewis, *Myanmar's top court hears Reuters reporters' appeal in official secrets case*, REUTERS (Mar. 25, 2019), https://www.reuters.com/article/us-myanmar-journalists/myanmars-top-court-hears-reuters-reporters-appeal-in-official-secrets-case-idUSKCN1R705B; Antoni Slodkowski & Simon Lewis, *Myanmar prosecutor seeks Official Secrets Act charges against two Reuters reporters*, REUTERS (Jan. 9, 2018), https://www.reuters.com/article/us-myanmar-journalists/myanmar-prosecutor-seeks-official-secrets-act-charges-against-two-reuters-reporters-idUSKBN1EY2S8.

[3] *Two Reuters journalists jailed: What is the Myanmar official secrets case?*, THE INDIAN EXPRESS (Sept. 3, 2018), https://indianexpress.com/article/what-is/what-is-the-myanmar-official-secrets-case/.

[4] Official Secrets Act § 3(1)(a) (Myan.)

[5] Official Secrets Act § 2(8) (Myan.)

[6] Official Secrets Act § 3(2) (Myan.)

[7] Official Secrets Act § 3(2) (Myan.); *"They Can Arrest You at Any Time": The Criminalization of Peaceful Expression in Burma*, HUMAN RIGHTS WATCH (June 29, 2016), https://www.hrw.org/report/2016/06/30/they-can-arrest-you-any-time/criminalization-peaceful-expression-burma.

By imprisoning journalists reporting on the oppression of the Rohingya, Myanmar created an informational vacuum which allowed the Burmese military to conduct an effective disinformation campaign on Facebook.[8] In this campaign, Burmese military officers created fake Facebook profiles, accumulated large followings through posts about Burmese celebrities, and then began to use these pages to disseminate false information about the Rohingya.[9] On these pages, the military posted images of corpses, claiming they were evidence of Rohingya perpetrated massacres, and called the Rohingya terrorists.[10] In 2017, the military used Facebook Messenger to warn Burmese Buddhists of imminent "jihad attacks" by Muslim groups.[11] At the same time, the military used the same tactic to warn the Rohingya of imminent anti-Muslim protests.[12] Through this misinformation campaign, the Burmese government inflamed ethnic tensions in the Northern Rakhine State, setting the stage for the ethnic cleansing that the military conducted soon thereafter.

B. Coup

While the Reuters journalists were released in 2019 after serving fewer than two years of their seven-year prison sentences,[13] Myanmar has continued to criminalize journalism, especially in the aftermath of the 2021 coup. Shortly after ousting the democratically-elected government in February, the Burmese military ordered journalists to stop describing their takeover with the words

[8] Paul Mozur, *A Genocide Incited on Facebook, With Posts From Myanmar's Military*, NEW YORK TIMES (Oct. 15, 2018), https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html.
[9] *Id.*
[10] *Id.*
[11] *Id.*
[12] *Id.*
[13] Simon Lewis & Shoon Naing, *Two Reuters reporters freed in Myanmar after more than 500 days in jail*, REUTERS (May 6, 2019), https://www.reuters.com/article/us-myanmar-journalists/two-reuters-reporters-freed-in-myanmar-after-more-than-500-days-in-jail-idUSKCN1SD056.

"coup," "regime," and "junta."[14] After many journalists failed to obey this command, the military cracked down on press freedoms, jailing at least 98 journalists by August.[15]

In order to more effectively suppress journalism following the coup, the military junta revised section 505(a) of the penal code.[16] The old version of the statute criminalized publishing any "statement, rumor or report…with intent to cause, or which is likely to cause, any officer, soldier, sailor or airman, in the Army, Navy or Air Force to mutiny or otherwise disregard or fail in his duty."[17] The new 505(a) uses even broader language, criminalizing "any attempt to hinder, disturb, damage the motivation, discipline, health and conduct of the military personnel and government employees and cause their hatred, disobedience, or disloyalty toward the military and the government."[18] This revised penal code has been widely used to imprison journalists in the country, and especially those who have described the military's takeover as a "coup."[19]

This criminalization of free reporting on the coup has strengthened the military's capacity to control the narrative around their government takeover.[20] Within this information vacuum, the military has claimed that it took power in response to a democratic emergency resulting from

[14] Richard C. Paddock, *Myanmar Soldiers, Aiming to Silence Protests, Target Journalists*, NEW YORK TIMES (Apr. 1, 2021), https://www.nytimes.com/2021/04/01/world/asia/myanmar-journalists-arrests.html.
[15] *Myanmar: Junta Escalates Media Crackdown*, HUMAN RIGHTS WATCH (July 27, 2021), https://www.hrw.org/news/2021/07/27/myanmar-junta-escalates-media-crackdown.
[16] *Myanmar: Post-Coup Legal Changes Erode Human Rights*, HUMAN RIGHTS WATCH (Mar. 2, 2021), https://www.hrw.org/news/2021/03/02/myanmar-post-coup-legal-changes-erode-human-rights.
[17] *Id.*
[18] *Myanmar Ruling Council Amends Treason, Sedition Laws to Protect Coup Makers*, THE IRRAWADDY (Feb. 16, 2021), https://www.irrawaddy.com/news/burma/myanmar-ruling-council-amends-treason-sedition-laws-protect-coup-makers.html; Myanmar Penal Code 505(a).
[19] Shawn W. Crispin, *Bitter reversal: Myanmar military coup wipes out press freedom gains*, COMMITTEE TO PROTECT JOURNALISTS (July 28, 2021), https://cpj.org/reports/2021/07/bitter-reversal-myanmar-journalists-jailed-imprisoned-military-crackdown/.
[20] Jenny Domino, *The Other De-Platforming We Should Have Been Talking About*, JUST SECURITY (May 11, 2021), https://www.justsecurity.org/76047/beyond-the-coup-in-myanmar-the-other-de-platforming-we-should-have-been-talking-about/.

alleged widespread voter fraud in the country's 2020 election.[21] Because of Covid-19 rules as well as its own sense of impending defeat, the Burmese military called for a postponement in the country's 2020 election, which was denied by Aung San Suu Kyi.[22] Given this context, the military claimed it was justified in taking power and that the election should be rerun in a year.[23]

This new wave of disinformation surrounding the coup has been relatively ineffective compared to the disinformation campaign the military conducted surrounding the Rohingya crisis. This relative ineffectiveness can be seen in the continuing public protests against the military government,[24] and might be explained by Facebook's increasing crackdown on disinformation on the platform as well as the closer relationship of the average Burmese citizen to the issue.

About three weeks after Myanmar's military coup, Facebook announced an immediate ban on military-linked pages.[25] Although this ban was delayed and it is unclear how effectively Facebook has been able to enforce it given the military's sophisticated and longstanding disinformation network on the platform, this move may have helped mitigate and reduce some of the spread of disinformation surrounding the coup, helping to prevent the military's narrative from controlling the site.

Alternatively, the disinformation campaign surrounding the military takeover may be relatively ineffective due to the tangible effect of the coup on the daily lives of the Burmese

---

[21] *Myanmar junta leader declares himself PM as election timeline stalled*, THE GUARDIAN (Aug. 1, 2021), https://www.theguardian.com/world/2021/aug/01/myanmars-military-ruler-promises-multi-party-elections.
[22] Ronan Lee, *COVID coup: how Myanmar's military used the pandemic to justify and enable its power grab*, THE CONVERSATON (Feb. 16, 2021), https://theconversation.com/covid-coup-how-myanmars-military-used-the-pandemic-to-justify-and-enable-its-power-grab-155350.
[23] *Id.*
[24] *More than 1,000 killed in Myanmar since February 1 coup*, AL JAZEERA (Aug. 18, 2021), https://www.aljazeera.com/news/2021/8/18/myanmar-coup-aapp-1000-killed-military.
[25] Paul Mozur, Mike Isaac, David E. Sanger & Richard C. Paddock, *COVID coup: how Myanmar's military used the pandemic to justify and enable its power grab*, NEW YORK TIMES (Feb. 24, 2021), https://www.nytimes.com/2021/02/24/technology/facebook-myanmar-ban.html.

citizens. In contrast to the coup, the Rohingya crisis was geographically limited to the relatively small and rural Northern Rakhine State on the outskirts of Myanmar and separated from much of the remainder of the country by the Arakan Mountains. Thus, the events taking place in that region were generally far removed from the daily lives of most Burmese people, making it more difficult for citizens to challenge the military narrative permeating the internet. By contrast, the coup played out across the entire country, allowing the Burmese citizens to better communicate about the situation and creating an informal, personal network where true information could spread throughout the country independent of news agencies and the internet.

II. Ethiopia

Since Ethiopia began an offensive against rebels in Tigray, the country has also cracked down on the freedom of the press, working to stifle reporting on the widespread atrocities that have occurred in the region.[26] When fighting broke out in Tigray in early November 2020, the Ethiopian government cut internet, mobile phone, and landline service to the region.[27] This information blackout made it difficult for journalists to receive and verify information about events taking place there.[28]

In addition to the media blackout in the region, within the first week of the conflict in Tigray, the Ethiopian government arrested at least six journalists.[29] Since that time, Ethiopia has

[26] Abdi Latif Dahir & Declan Walsh, *As Ethiopia Fights in Tigray Region, a Crackdown on Journalists*, NEW YORK TIMES (May 13, 2021), https://www.nytimes.com/2021/05/13/world/africa/ethiopia-tigray-journalists.html.
[27] Salem Solomon, *Journalists Struggle Through Information Blackout in Ethiopia*, VOICE OF AMERICA NEWS (Dec. 4, 2020), https://www.voanews.com/a/press-freedom_journalists-struggle-through-information-blackout-ethiopia/6199045.html.
[28] *Id.*
[29] *Id.*

continued to arrest journalists who have reported on the ongoing conflict.[30] The Ethiopian government has argued that these arrests are not based on the journalists' profession, but "due to their affiliation with a terrorist group which is banned by the parliament."[31] The terrorist group referred to in this statement is the Tigray People's Liberation Front, and this designation and the resulting criminalization of association with the group likely has a chilling effect on journalists reporting on the conflict.[32]

These journalists in Ethiopia have largely been charged with violating Ethiopia's Anti-Terrorism Proclamation.[33] Under this law, individuals who publish statements considered to be encouraging terrorism can be punished with ten to twenty years in prison.[34] This has been used to criminalize journalists who criticize government actions. Since the beginning of the conflict in Tigray, this has been used especially to punish individuals who have published articles critical of the Ethiopian government or sympathetic to the Tigray People's Liberation Front, since this group has been designated a terrorist organization.

In addition to using the Anti-Terrorism Proclamation to criminalize journalism, the Ethiopian government has also systematically revoked the permits of journalists who have spoken

[30] *Ethiopian authorities arrest at least 15 employees of 2 independent media outlets*, COMMITTEE TO PROTECT JOURNALISTS (July 2, 2021), https://cpj.org/2021/07/ethiopian-authorities-arrest-at-least-15-employees-of-2-independent-media-outlets/.
[31] *Id.*
[32] Mengesha Amare, *Ethiopia: TPLF, Shene to Be Designated As Terrorist Organizations,* THE ETHIOPIAN HERALD (May 2, 2021), https://allafrica.com/stories/202105030332.html.
[33] Lindsay Church, *Striking the Balance: Combating Terrorism and Preserving the Freedom of Expression in Ethiopia,* https://harvardilj.org/2016/01/striking-the-balance-combating-terrorism-and-preserving-the-freedom-of-expression-in-ethiopia/.
[34] *Id.*

out about the crisis in Tigray.[35] Foreign journalists reporting on the crisis are also regularly deported from the country.[36]

This crackdown on free media coverage of the conflict has allowed the Ethiopian government and individuals on social media to largely monopolize reporting on the situation. Additionally, where non-state-controlled media continues to report on the situation, media blackouts in the region have prevented them from fully verifying facts about the conflict, increasing their reliance on manipulated and inaccurate sources.[37] As a result, there have been numerous instances of false information being spread, including manipulated images taken from other conflicts and widespread fake Twitter accounts.[38]

Within this informational vacuum, the Ethiopian government has pushed its own narrative about the events in the region, describing its actions there as a "law and order operation" in response to Tigrayan aggression.[39] Furthermore, the Ethiopian government has consistently discounted and obscured reports of human rights abuses and war crimes occurring in the region.[40] Despite United Nations reports indicating that only twelve percent of Tigrayans have received necessary humanitarian aid, the government has argued that it has reached nearly all of those in need.[41] These government narratives have been reported on fake Twitter accounts, including one

---

[35] Meron Gebreananaye, *Ethiopia: TPLF, Shene to Be Designated As Terrorist Organizations,* THE ETHIOPIAN HERALD (June 23, 2021), https://www.ethiopia-insight.com/2021/06/23/hands-off-ethiopia-a-new-phase-in-the-tigray-disinformation-campaign/.
[36] *Ethiopia Expels Crisis Group Senior Analyst,* INTERNATIONAL CRISIS CROUP (Nov. 22, 2021), https://www.crisisgroup.org/africa/horn-africa/ethiopia/ethiopia-expels-crisis-group-senior-analyst.
[37] *Disinformation in Tigray: Manufacturing Consent for a Secessionist War*, New Africa Inst. (May 2021),https://www.scribd.com/document/507224143/Disinformation-in-Tigray-Manufacturing-Consent-For-a-Secessionist-War#from_embed.
[38] Peter Mwai, *Tigray conflict: The fake UN diplomat and other misleading stories,* BBC NEWS (Mar. 25, 2021), https://www.bbc.com/news/56456535.
[39] Meron Gebreananaye, *"Hands Off Ethiopia": A new phase in the Tigray disinformation campaign,* ETHIOPIA INSIGHT (June 23, 2021), https://www.ethiopia-insight.com/2021/06/23/hands-off-ethiopia-a-new-phase-in-the-tigray-disinformation-campaign/.
[40] *Id.*
[41] *Id.*

posing as a former UN diplomat, which was used to criticize calls for investigations into atrocities in the region.[42]

III. North Korea

Despite asserting a constitutional right to freedom of the press, North Korea has perhaps the most strictly-state controlled media in the world. In North Korea, radios and televisions are pre-tuned to receive only government stations.[43] It is a crime for individuals to listen to foreign media, or even to own a radio or television capable of being tuned to receive stations other than the North Korean state media.[44] North Korean citizens are also blocked from the international internet, receiving only a tightly-controlled intranet.[45] Furthermore, cell phones in the country are unable to receive international calls and cell phone usage is strictly monitored by security forces.[46] Individuals who smuggle cell phones from abroad are often sent to reeducation camps.[47] Since Kim Jong-un took power, the country has been even more aggressive in digital monitoring and has imposed even harsher penalties for those found in violation of media restrictions.[48] In addition to programs which automatically delete unapproved files from devices connected to the North Korean intranet, North Korean security personnel in "Group 109" are charged with physically inspecting North Korean citizens' devices in order to ensure that they contain no foreign media.[49] This is an important additional measure because, due to the extreme limitations of the North

[42] *Id.*
[43] *North Korea's tightly controlled media,* BBC NEWS (Dec. 19, 2011), https://www.bbc.com/news/world-asia-pacific-16255126.
[44] Robert R. King, *North Koreans Want External Information, But Kim Jong-Un Seeks to Limit Access,* CENTER FOR STRATEGIC & INTERNATIONAL STUDIES (May 15, 2019), https://www.csis.org/analysis/north-koreans-want-external-information-kim-jong-un-seeks-limit-access.
[45] *Id.*
[46] *Id.*
[47] *Id.*
[48] *Id.*
[49] *Id.*

Korean intranet, many North Koreans rarely connect their personal devices to the network.[50] Thus, in the absence of this secondary measure, these individuals may be able to keep their illicit files by simply not using the North Korean network.[51]

When foreign journalists are permitted to enter the country, they are accompanied at all times by government minders to ensure that they receive a carefully curated and state-approved image of North Korean life.[52] Journalists who are found to be critical of North Korea while visiting the state have been detained and held for interrogation.[53] When they are permitted to interview citizens, they are given carefully-selected citizens who dutifully report the state line.[54] This makes it nearly impossible for international news sources to develop and report an accurate depiction of everyday life in North Korea. Because there is so little real reporting out of North Korea, the outside world must rely to some extent on these curated depictions by the government. Thus, despite journalists' awareness that they are being presented a carefully crafted and manipulated image of North Korean life, they are left to report on the experience as it occurred, even while they work to acknowledge the performative elements of their experiences and visits. Through these restrictions, these government-crafted depictions become more central to the outside world's view of North Korea, since there is so little other information leaving the nation's borders.

North Korea also aggressively punishes journalists who criticize the government. In 2017, for example, a North Korean court sentenced two South Korean journalists to death for positively

---

[50] *Id.*
[51] *Id.*
[52] Maeve Shearlaw, *Propaganda and the party congress: how to report from North Korea,* THE GUARDIAN (May 5, 2016), https://www.theguardian.com/world/2016/may/05/propaganda-and-the-party-congress-how-to-report-from-north-korea.
[53] Phil Robertson, *Dispatches: North Korea's Idea of "Freedom of Expression",* HUMAN RIGHTS WATCH (May 11, 2016), https://www.hrw.org/news/2016/05/11/dispatches-north-koreas-idea-freedom-expression.
[54] Robert R. King, *North Koreans Want External Information, But Kim Jong-Un Seeks to Limit Access,* CENTER FOR STRATEGIC & INTERNATIONAL STUDIES (May 15, 2019), https://www.csis.org/analysis/north-koreans-want-external-information-kim-jong-un-seeks-limit-access.

reviewing a book describing an increasing market economy in North Korea.[55] Such harsh penalties have also been given to journalists who have entered the country outside of the prescribed system of journalist access. For example, in 2009, two American journalists were sentenced to twelve years of hard labor for illegally entering the country to report on North Koreans crossing the Tumen River to enter China.[56]

By criminalizing journalism to such a high degree, North Korea has created a system in which nearly all information available to its citizens is within the exclusive control of the government. This allows the North Korean government to effectively indoctrinate its citizens through an abundance of carefully curated propaganda celebrating every governmental success and minimizing and manipulating the country's shortcomings.[57] These techniques are particularly important in an authoritarian regime like North Korea, where they help to incentivize citizens to maintain loyalty to the regime and to continue to provide labor to support the regime despite the relatively limited benefits they receive from this work.[58] Because of the strict and complete control the North Korean government has claimed over media and journalism, the country has been particularly effective in pushing its disinformation campaigns to its citizens.

Furthermore, by strictly limiting access of foreign journalists to the territory and by preventing citizens from having access to cross-border communications, North Korea's strict

[55] *North Korea sentences South Korean reporters to death over review of book about country,* REUTERS (Aug. 31, 2017), https://www.reuters.com/article/us-northkorea-southkorea-media-threat/north-korea-sentences-south-korean-reporters-to-death-over-review-of-book-about-country-idUSKCN1BB2J0.
[56] Justin McCurry, *North Korea sentences two US journalists to 12 years in jail,* THE GUARDIAN (June 8, 2009), https://www.theguardian.com/world/2009/jun/08/north-korea-us-journalists.
[57] Tae-jun Kang, *North Korea Strengthens Propaganda Efforts Ahead of Key Party Anniversary,* THE DIPLOMAT (Oct. 9, 2020), https://thediplomat.com/2020/10/north-korea-strengthens-propaganda-efforts-ahead-of-key-party-anniversary/.
[58] Hyung-Jin Kim, *North Korea waging propaganda-heavy, 80-day labor campaign,* ASSOCIATED PRESS (Nov. 13, 2020), https://apnews.com/article/international-news-seoul-south-korea-north-korea-coronavirus-pandemic-ffe76e53a73f1879c149cb9954d95e08.

system of media control has also prevented foreigners from receiving much news directly from the country, leaving news stations to rely upon testimony from defectors and remote investigation methods, such as satellite imagery. This informational vacuum was recently observed in the context of the Covid-19 pandemic. North Korea claims it has not had any infections.[59] This claim is doubted by the United Nations, especially due to the aggressive containment measures North Korea has taken, including closing its border with China, at the expense of its already fragile economy.[60] However, it has been difficult for the outside world to estimate the virus's effect on the nation due to the lack of reporting leaving the country's borders.[61]

Conclusion

Through these and other methods, countries have criminalized journalism and created media vacuums. These media vacuums ensure that the state has a greater control on the narrative being presented to the world by eliminating other sources which may fact check the government or provide alternative perspectives on a situation. Thus, criminalization and control of journalism and free press allows state disinformation campaigns in such states to be particularly effective.

[59] *Coronavirus: How the pandemic is hitting North Korea hard,* DEUTSCHE WELLE https://www.dw.com/en/coronavirus-how-the-pandemic-is-hitting-north-korea-hard/a-57168554.
[60] Mitch Shin, *What Is the Truth About COVID-19 in North Korea?,* THE DIPLOMAT (Jan. 6, 2021), https://thediplomat.com/2021/01/what-is-the-truth-about-covid-19-in-north-korea/.
[61] *Coronavirus: How the pandemic is hitting North Korea hard,* DEUTSCHE WELLE https://www.dw.com/en/coronavirus-how-the-pandemic-is-hitting-north-korea-hard/a-57168554.