Stanford Law School
Law and Policy Lab

How to Reconcile International Human Rights Law and Criminalization of Online Speech: Violent Extremism, Misinformation, Defamation, and Cyberharassment

2019-2020 PRACTICUM RESEARCH TEAM

Amélie-Sophie VAVROVSKY
Anirudh JAIN
Asaf ZILBERFARB
David JAFFE
Eric FRANKEL
Jasmine SHAO
June LEE
Justin WONG
Madeline LIBBEY
Madeline MAGNUSON
Naz GOCEK
Nil Sifre TOMAS
Shalini IYENGAR
Sydney FRANKENBERG

INSTRUCTORS AND PROJECT LEADS:

Sarah SHIRAZYAN, Ph.D.
Lecturer in Law

Allen WEINER
Senior Lecturer in Law,
Director, Stanford Program in International and Comparative Law

Yvonne LEE, M.A.
Teaching Assistant

Madeline Magnuson, J.D.
Research Assistant

ABOUT THE STANFORD LAW SCHOOL POLICY LAB

Engagement in public policy is a core mission of teaching and research at Stanford Law School (SLS). The Law and Policy Lab (The Policy Lab) offers students an immersive experience in finding solutions to some of the world's most pressing issues. Under the guidance of seasoned faculty advisers, Policy Lab students counsel real-world clients in an array of areas, including education, global governance, transnational law enforcement, intellectual property, policing and technology, and energy policy.

Policy labs address policy problems for real clients, using analytic approaches that supplement traditional legal analysis. The clients may be local, state, federal and international public agencies or officials, or private non-profit entities such as NGOs and foundations. Typically, policy labs assist clients in deciding whether and how qualitative and/or quantitative empirical evidence can be brought to bear to better understand the nature or magnitude of their particular policy problem and identify and assess policy options. The methods may include comparative case studies, population surveys, stakeholder interviews, experimental methods, program evaluation or big data science, and a mix of qualitative and quantitative analysis. Faculty and students may apply theoretical perspectives from cognitive and social psychology, decision theory, economics, organizational behavior, political science or other behavioral science disciplines. The resulting deliverables reflect the needs of the client, with most resulting in an oral or written policy briefing for key decision-makers.

Directed by former SLS Dean Paul Brest, the Policy Lab reflects the school's belief that systematic examination of societal problems, informed by rigorous data analysis, can generate solutions to society's most challenging public problems. In addition to policy analysis, students hone the communications skills needed to translate their findings into actionable measures for policy leaders and the communities they serve. The projects emphasize teamwork and collaboration and are often interdisciplinary, giving law students the opportunity to work with faculty and colleagues from across the university with expertise in such fields as technology, computer science, medicine, business and international diplomacy, among others.

ABOUT THE STANFORD LAW SCHOOL POLICY LAB

Engagement in public policy is a core mission of teaching and research at Stanford Law School (SLS). The Law and Policy Lab (The Policy Lab) offers students an immersive experience in finding solutions to some of the world's most pressing issues. Under the guidance of seasoned faculty advisers, Policy Lab students counsel real-world clients in an array of areas, including education, global governance, transnational law enforcement, intellectual property, policing and technology, and energy policy.

Policy labs address policy problems for real clients, using analytic approaches that supplement traditional legal analysis. The clients may be local, state, federal and international public agencies or officials, or private non-profit entities such as NGOs and foundations. Typically, policy labs assist clients in deciding whether and how qualitative and/or quantitative empirical evidence can be brought to bear to better understand the nature or magnitude of their particular policy problem and identify and assess policy options. The methods may include comparative case studies, population surveys, stakeholder interviews, experimental methods, program evaluation or big data science, and a mix of qualitative and quantitative analysis. Faculty and students may apply theoretical perspectives from cognitive and social psychology, decision theory, economics, organizational behavior, political science or other behavioral science disciplines. The resulting deliverables reflect the needs of the client, with most resulting in an oral or written policy briefing for key decision-makers.

Directed by former SLS Dean Paul Brest, the Policy Lab reflects the school's belief that systematic examination of societal problems, informed by rigorous data analysis, can generate solutions to society's most challenging public problems. In addition to policy analysis, students hone the communications skills needed to translate their findings into actionable measures for policy leaders and the communities they serve. The projects emphasize teamwork and collaboration and are often interdisciplinary, giving law students the opportunity to work with faculty and colleagues from across the university with expertise in such fields as technology, computer science, medicine, business and international diplomacy, among others.

TABLE OF CONTENTS

I. EXECUTIVE SUMMARY

The rights to freedom of speech and expression are considered to be at the heart of international human rights law protections. In practice, however, this right has frequently been in tension with other rights protected by human rights law, such as the right to reputation and the right against discrimination. Indeed, in recognition of the dangers of untrammeled freedom, international human rights law also provides for limitations on the exercise of these rights. With the rise of the internet, laws surrounding freedom of speech and expression have had to confront new ambiguities, controversies, and tensions. While previous revolutions in transmission such as the printing press have fundamentally impacted speech and association in the past, the reach, velocity, immediacy, anonymized and decentralized nature of information transmitted on the internet raises unique free speech and law enforcement challenges.

The proliferation of online communication and social media activity as dominant modes of expression necessitate an examination of that new developments in international human rights law in four broad areas: (1) online manifestations of support for conduct by violent extremist organizations; (2) misinformation and fake news; (3) online defamation; and (4) cyber harassment and bullying. Our research analyzes whether and the extent to which online activities in these four key areas are: (1) protected by international human rights law, as analyzed in jurisprudence; (2) treated as crimes under domestic criminal law; and (3) banned under selected content policies of leading social media companies. Accordingly, our observations draw from legal research on human rights treaties, past decisions by domestic and international courts, domestic legal codes, and other academic articles. This research was primarily restricted to source documents available online in English.

This report begins with an introductory chapter on the architecture of international human rights law on free expression and association, identifying the key conventions that enshrine these rights and interpretative guidance issued by international monitoring bodies and conferences of legal experts. We articulate the jurisprudential principles of freedom of expression common across our four subject matter areas. First, these freedoms apply to online activities as well as to the offline world. Second, in order for restrictions on speech or association not to violate international human

rights law, the restriction must pass a three-part test. The restriction must (1) be prescribed by law, (2) pursue one of a specified list of legitimate aims, including but not limited to public order and safety, and (3) be necessary in a democratic society as well as proportionate.

The report next analyzes our four subject-matter areas: (1) violent extremist organizations (VEOs); (2) misinformation and fake news; (3) online defamation; and (4) cyber harassment and bullying. Each of these four chapters analyzes trends in international jurisprudence, national legislation, and social media content policies.

A. Violent Extremist Organizations Online: Expression and Association

This chapter addresses online expression and association relating to violent extremist organizations by: outlining the range of VEO activities online, synthesizing international jurisprudence on prosecutions of defendants for praising, supporting, and representing VEOs, surveying national legislation criminalizing these activities, and characterizing social media platform policies relating to extremist content.

First, the chapter outlines the different activities and functions of terrorist and other extremist groups on social media, including audience development, incitement, information collecting, financing, and online disruption. In order to promote their respective causes, violent extremist organizations (VEOs) engage in a large range of activities online, spanning from activities that clearly constitute ordinary crimes—such as hacking or threatening violent attacks—to activities that squarely implicate freedom of expression concerns.

Second, the chapter examines jurisprudence from international and national courts analyzing whether criminalizing different categories of extremist activities complies with or violates international rights to freedom of expression and association. This research is organized according to categories of extremist activity online: representation, operational support, and ideological support. Cases on representation involved criminal prosecutions of people for declaring themselves supporters of, otherwise displaying membership in, or performing recruitment on behalf of, entities engaging in illegal violence. Cases on operational support activities involved defendants collecting and transferring funds on behalf of, or coordinating and

planning attacks for, such entities. Cases on ideological support involved defendants sharing or reposting an entity's propaganda without commentary indicating sentiment, encouraging future attacks, praising past attacks, and praising the entity's goals, ideology, leaders, or members.

In total, the research for this section identified twenty-nine cases that examined issues of freedom of expression or association related to VEO activity, whether online or offline. The vast majority of these cases related to ideological support. In almost all of these cases, courts consistently articulated the three-part test of prescription by law, pursuit of legitimate aims, and necessity and proportionality. The substance of the courts' analysis almost always lay in analyzing the necessity and proportionality of the restriction.

Accordingly, the research team developed an analytical framework of eight factors for assessing necessity and proportionality: (1) nature of the post; (2) number of the post(s); (3) content of the post; (4) resulting violence, if any; (5) timing of the post; (6) medium and reach; (7) speaker's role and personal history, and (8) proportionality of the sentence. This analytical framework is described in more detail in the Key Takeaways section. Three categories of content never, in the cases we analyzed, justified criminal prosecution resulting in sentences of imprisonment: (1) criticizing the government's suppression of a VEO; (2) praising a VEO's goals or ideology without endorsing its violent methods; and (3) praising a VEO's leader without referencing his violent or criminal acts.

Third, the chapter identifies four emerging trends in different countries' domestic laws addressing dangerous groups' adaptations to social media. This section analyzes eight countries and their laws regulating and criminalizing extremist speech and speech supporting terrorist organizations: Germany, Pakistan, Russia, Australia, Brazil, Kenya, Singapore, and Egypt. Some of these countries rely on legislation designed for regulating "offline" speech as the primary vehicle for regulating online speech, but others have implemented extremism laws specifically designed to address online activity. The surveyed countries use a mixture of laws to address both domestic extremist groups as well as jihadist terrorist organizations; many passed jihadist-focused laws in the wake of the September 11th attacks, but a few have also passed legislation specifically targeting right-wing speech online. Many of the countries' laws had vague definitions of terrorism, and others had extremely vague definitions of prohibited incitement to terrorism, opening the door

to prosecution based on speech that would not generally be considered incitement. Two countries impose individual criminal liability not only on people who post extremist content online but also intermediary hosts who fail to remove such content.

Fourth, the chapter analyzes the policies adopted by social media companies related to VEO content. Every social media platform we assessed has a policy that prohibits terrorist organizations from using its forums to recruit members or advocate for violence. Every platform extends this prohibition to re-posting terrorist propaganda or sharing content that glorifies terrorist leaders or violent acts or posting terrorist symbols or insignia in a positive light. Some, but not all, make explicit exceptions for educational or awareness-raising purposes, but place the burden on the poster to make this context clear. Most platform policies reflect that context and intent are essential factors to consider when contemplating removing posts and terminating accounts.

B. Spreading False Information Online

This chapter examines ongoing efforts to criminalize the spread of false information online. The term of false information, as used in this chapter, encompasses both misinformation (the spread of *unintentionally* false information) and disinformation (the spread of *intentionally* false information). The chapter proceeds by: analyzing the implications of jurisprudence for criminalizing the spread of false information; quantitatively and qualitatively characterizing global and regional trends in state responses to misinformation and disinformation; sampling how different countries have criminalized the spread of false information and how national courts have responded; and presenting social media policies related to misinformation and disinformation.

First, the chapter analyzes the extent to which criminalization of spreading false information complies with international law as represented in international jurisprudence. In four key areas, spreading particular types of false information falls under existing legal frameworks for expression: incitement to violence or hate speech, fraud or false advertising, defamation, and 'memory' laws concerning false information about historical occurrences. International human rights law jurisprudence clearly permits the criminalization of false information that rises to the

level of fraud, incitement to violence, or hate speech, but generally prohibits the criminalization of mere defamation, as discussed in more detail in the chapter on defamation.

Turning to the criminalization of spreading false information outside these four key areas, established international legal principles yield less clarity. Many international organizations agree that general prohibitions on the dissemination of information based on "vague and ambiguous ideas," such as "false news," are "incompatible with human rights law and should be abolished."[1] The relatively few international cases on point indicate the classic three-part test of legal prohibition, necessity, and proportionality still applies to any restriction on the spread of generalized false information. International courts applying human rights treaties have not applied this test to hold that the criminalization of misinformation or disinformation *per se* violates the right to freedom of expression contained in their respective treaties. However, they have held in particular instances that prohibitions on spreading false information were too vague, or that the speech at issue—being political speech—was too important to bear criminal restriction.

Second, the chapter examines state practice on criminalizing the spread of false information, via quantitative findings, qualitative trends observed, and a sampling of measures taken by specific countries. This chapter draws on a survey of almost all countries in the world—194 countries—organized according to geographic and political region. Legislative responses to the spread of false information include criminalization, imposition of civil intermediary liability, and awareness campaigns.

The quantitative analysis reveals both global and regional trends. Globally, countries that have criminalized or at least legislatively considered criminalizing misinformation or disinformation constitute, overall, a minority of countries. However, they make up a majority of the countries that have legislatively responded to the spread of false information. Additionally, statistical analysis characterizes which regions legislate with regard to misinformation and disinformation in an atypical fashion. The COVID-19 pandemic has sparked about a quarter of all

[1] *Joint Declaration on Freedom of Expression and Fake News, Disinformation and Propaganda*, The United Nations (UN), the Organization for Security and Co-operation in Europe (OSCE), the Organization of American States (OAS) the African Commission on Human and Peoples' Rights (ACHPR), 2017. Available at: https://www.osce.org/fom/302796?download=true

actions proposed or taken against false information. The majority of those pandemic-related actions have included criminal penalties.

A qualitative summary of trends observed is followed by a sampling of specific laws passed in different regions of the world as well as how national courts have responded to the question of whether or not criminalization of sharing of false information is compatible with the right to freedom of expression. The most common trend seemed to be that decisions to prosecute people for spreading false information are generally contingent on the impact of the dissemination, and laws criminalizing the spread of false information tend to focus on the consequences of the action (such as a detriment to public health) rather than on the action itself. National courts have considered whether the concept of "false information" is too vague to be used in a criminal statute; their inquiries have produced mixed outcomes.

Third, the chapter reveals the substantial similarity between YouTube, Twitter, and Facebook's policies relevant to misinformation and disinformation. These three platforms all ban posts of substantially manipulated media, fake accounts that impersonate others in misleading ways, spreading false information tending to suppress voting or census participation, and use of multiple accounts in ways that artificially manipulate conversations or mislead users in specified ways. Facebook also generally bans misinformation creating a risk of imminent violence or physical harm. This chapter includes an analysis of how these platforms have responded to the COVID-19 pandemic.

C. Defamation Online

This chapter analyzes the tensions inherent in defamation law, seeking to balance international human rights to free expression and reputation online. This analysis, of how restrictions on defamatory speech alternately comply and violate the international right to freedom of expression, proceeds by: presenting trends in international legal jurisprudence on defamation, surveying national laws and identifying regional legislative trends, and summarizing the approaches taken to defamation by social media platforms.

First, this chapter briefly indicates which treaties enshrine the rights to reputation and privacy as principles of international human rights law, then analyze trends in international legal jurisprudence on defamation. These include: (i) that to be defamatory, statements must be factual rather than opinion-based; (ii) that despite state practice to the contrary, international courts consider criminalization of defamation to be too severe a form of restriction to comply with free expression, (iii) that free expression requires governments to show particular restraint in restricting criticism of public figures and heads of state, and (iv) that disagreement exists over the permissibility of criminal blasphemy laws.

Second, the chapter surveys national laws and identify regional legislative trends in Europe, the Americas, the Middle East, Africa, and the Asia Pacific. For each region, this section traces how state practice reflects and informs the abovementioned issues, including *desacato* or *lèse majésté* laws imposing heightened penalties on defamation criticizing public officials or heads of state. Criminal defamation laws were generally drafted with offline application in mind, but are applied to online speech as well. Criminal defamation laws are widespread in Europe and the Asia Pacific region, in the latter of which *lèse majésté* laws are particularly common. A number of Middle Eastern countries have and vigorously enforce criminal defamation laws. Some countries in Africa and the Americas also have criminal defamation laws. However, countries in the Americas appear to be moving away from criminalizing defamation; at least eleven such countries have lifted criminal liability for defamation.

Third, the chapter summarizes the approaches taken to defamation by social media platforms. Most major platforms do not mention defamation in their community guidelines; instead, they process legal take-down requests on the basis of local illegality. For instance, if a defamed person submits a court order declaring content illegal in a particular country, platforms will generally block access to that content in that country.

D. Cyberharassment and Cyberbullying

This chapter presents an overview of national and international law and jurisprudence related to some of the main forms of cyberharassment. This is an umbrella term defined in this

report as encompassing all harassment that takes place online, including behaviors such as cyberstalking, online harassment of minors (which we refer to as "cyberbullying"), the dissemination of revenge porn, and other harmful online conduct ("other cyber harms"). This analysis proceeds by: indicating which international conventions are relevant to cyberharassment and hate speech; analyzing in turn the relevant national jurisprudence and state practice for cyberstalking, harassment of minors, sexual harassment, and other cyber harms including doxing; and then summarizing social media platform policies on hateful conduct.

First, this chapter briefly indicates which international human rights law (IHRL) treaties are relevant to cyberharassment. No international or regional treaties directly address cyberharassment. However, there are some provisions that address and encourage states to prohibit harassment and cyberharassment within particular domains, namely gender and the workplace. For context, this section also points to treaty provisions on hate speech.

Second, this chapter surveys any relevant principles of international jurisprudence, trends in national legislation, and notable court cases for each of the following categories of cyberharassment: cyberstalking, harassment of minors, sexual harassment, and other cyber harms including doxing (posting a target's private information online, such as his phone number or credit card information). Most on-point court cases arose in national courts rather than international bodies, and few found any violation of the right to free expression.

Cyberstalking generally entails repeated or persistent threatening conduct online with harmful effects for another. Numerous countries consider cyber stalking to be one of the more serious forms of harassment. This conduct is frequently criminalized in national legislation, although the harshness of the penalty varies across countries.

Cyberbullying, defined in this report as the online harassment of minors, is generally not criminalized. Instead, numerous countries have passed or amended laws to address cyberbullying in the context of the education system, often in the wake of high-profile tragic events involving young victims. Cyberbullying undertaken by minors is generally not criminalized, except in some U.S. states. Instead, most of these laws charge educational institutions with addressing cyberbullying, while noting that if the issue is serious enough (usually, if it independently violates the criminal code) the police may get involved.

Sexual harassment online includes "revenge porn," which generally entails the online distribution of someone's intimate images or videos without consent, often with the intent of causing embarrassment or distress. The online distribution of sexual images without consent and similar actions that fundamentally violate someone's privacy and bodily autonomy tend to be criminalized, though the harshness of the penalty varies across countries.

Other cyber harms are covered in a residual category that we describe as online conduct resulting in harm, whether physical, mental, psychological, or economic. Countries are most likely to criminalize conduct that facilitates violence or physical damage, which sometimes falls under incitement to violence. However, countries also impose criminal sanction for other harmful online behaviors that result in fear, alarm, annoyance, or distress. Many but not all countries require that the perpetrator possess an intent to cause harm or knowledge that their behavior would cause harm. The publication of identity information that can cause similar harm (i.e. doxing) is a particular subset of this conduct that has been criminalized in multiple countries.

Third, this chapter summarizes social media platform policies on hateful conduct. These platforms' community guidelines generally prohibit users from engaging in hate speech, threats of violence or property damage, and cyberharassment or cyberbullying, as well as doxing of private personally identifiable information.

Conclusions on legislative and jurisprudential trends across categories of cyberharassment follow. In terms of state practice, cyberharassment activities tend to be at least lightly criminalized, i.e. eligible for statutory penalties of less than two years in prison. The exception is online harassment involving minors, which is only specifically criminalized in a number of U.S. states; most countries have concluded cyberbullying is best handled within the education system. A number of countries have yet to create laws that specifically target online manifestations of these issues. However, even these countries generally have "offline" laws which can be applied to prosecute harassment online.

Jurisprudentially, our research unearthed only three instances of criminal responsibility for cyberharassment being held to violate freedom of expression: in all three cases, the speech at issue was politically charged. In two of these cases, the speech for which defendants were prosecuted was directed against government officials. One of these was decided under international principles

by the OHCR's Working Group on Arbitrary Detention, in *Nyanzi v. Uganda*. Two were decided by national courts under constitutional principles: the *Singhal* case in India and *Elliott* case in Canada.

II. KEY TAKEAWAYS

This section intends to highlight high-level takeaways that may be of particular importance and represent guiding principles to inform analysis of whether or not a particular restriction on speech is compliant with international human rights law. Notable takeaways from this report include:

1. Online speech related to Violent Extremist Organizations (VEOs)
    a. An analytical framework of 8 key factors used to assess a restriction's necessity and proportionality, derived from cases analyzing prosecutions for praising, supporting, or representing VEOs.
2. Spreading false information online: jurisprudential and national trends
    b. Four relevant established frameworks relate to false information in international jurisprudence (fraud, incitement to violence, hate speech, and defamation); these frameworks clearly permit criminalizing false information that rises to the level of fraud, incitement to violence, or hate speech, but generally prohibits the criminalization of mere defamation.
    c. Outside of these established frameworks, the relatively few international cases on point indicate the classic three-part test of legal prohibition, necessity, and proportionality still applies to any restriction on the spread of generalized false information, and have held that particular prohibitions on spreading false information were too vague to support restriction.
    d. Quantitative observations on global trends with respect to legislation combatting and criminalizing the spread of false information. Globally, about one-third of countries have criminalized spreading false information.
3. Defamatory speech online: jurisprudential and national trends
    e. Facts versus Opinion: To be defamatory, statements should be factual rather than opinion based.

f. Criminalization of Defamation: Despite state practice to the contrary, international courts consider criminalizing defamation too severe a form of restriction to comply with free expression.
g. Criticism of Public Officials: Free expression requires states to show particular restraint in restricting criticism of public figures and heads of state.
h. Defamation of Religion: Application of criminal blasphemy laws may comply with free expression only when the speech in question incites discrimination, hostility, or violence.

4. Cyberbullying and cyberharassment: jurisprudential and national trends
    i. Few court cases found applications of cyberharassment laws to violate free expression.
    j. Many forms of cyberharassment are at least lightly criminalized, including cyberstalking, sexual harassment, and doxing.
    k. Cyberbullying by minors is generally not criminalized.

A. VEO-Related Speech: Analytical Framework for Assessing the Necessity and Proportionality of Restrictions on Speech that Purportedly Endorses Violence

Based on twenty-nine cases analyzing restrictions related to VEOs, the VEO-specific research team developed a framework of eight key factors that courts consider when determining whether or not a restriction on free speech is justified, i.e. whether or not it is necessary in a democratic society and proportionate. None of these cases analyzed all of these factors, but they all mentioned at least two of these factors.

Glorification or praise of violence may constitute incitement, and therefore justify restrictions on free expression (including criminalization), when the content endorses violence. These factors provide a way of analyzing whether a restriction is justified based on an allegation that the content endorses violence. This framework resembles guidelines like the Rabat Plan of Action and Johannesburg Principles, but is not derived from them. Rather, all these typologies of

factors follow the inherent logic of how to weigh both the content and the context of a prosecuted post.

The eight factors fall into the following four categories:

A. The post itself
    1) Nature of the post
    2) Number of the post(s)
    3) Content of the post

B. Impact of the post
    4) Resulting violence, if any

C. Context of the post
    5) Timing of the post
    6) Medium and reach
    7) Speaker's role and personal history

D. Severity of restriction based on the post
    8) Proportionality of the sentence

In practice, courts varied in how they weighed these elements against each other. Many courts, in order to uphold criminalization of speech, required the content of the post to explicitly advocate violence—even when contextual factors weighed heavily in favor of restriction.[2] However, other courts were willing to uphold criminalization even if a post had only implicitly endorsed violence (such as by ambiguous but mildly positive references to former attacks). These courts weighed contextual factors more heavily than content-based ones.[3]

The function of each factor is explained below:

[2] E.g., *Hussain v. The Norwegian Prosecution Authority*, Nos. 14-0499903AST-BORG/01, 14-174730AST-BORG/01 (Norwegian App. Ct. 2015).
[3] E.g., *Zana v. Turkey*, No. 40984/07 (ECtHR Grand Chamber 1997).

1. Nature of the Post(s)

Courts consider whether a post is intended or appears to be satire rather than sincere advocacy,[4] or is phrased hypothetically.[5,6] Even a clear satirical or joking stance, however, does not always save a post from being considered to justify restriction on free expression.[7]

2. Number of the Post(s)

Courts also consider the number of complained-of posts. If a speaker or entity has engaged in a large number of provocative statements,[8,9,10] courts are more likely to ascribe intent to endorse or encourage violence than if the statements complained of are few in number.[11]

3. Content of the Post

The single most important factor—and sometimes the only one to which courts pay attention, other than proportionality of the sentence—is what the post says. Courts consider whether the post is specific or ambiguous in its support for a VEO or its violent activities.

When only one remark within a lengthy speech or post appears to support violent extremist activity, courts often consider the text surrounding the remark and text found elsewhere in the piece. For instance, courts ask whether the sentences alleged to incite violence are aberrant in the larger context of the piece or mediated by more pacific sentences elsewhere in the post, versus if they appear to represent the post's central message or thesis.[12]

There is a spectrum of likelihood that courts will find a restriction, especially criminalization, justified based on the content of a piece of speech. Justification is very likely for content that explicitly encourages violence,[13] compared to moderately likely for content that

4 *The State v. Cassandra Vera*, STC 493/2018 (Spanish Supreme Court 2018).
5 *Smajić v. Bosnia and Herzegovina*, No. 48657/2016,(ECtHR 2018).
6 *Fatullayev v. Azerbaijan*, No. 40984/07 (ECtHR 2010).
7 *Leroy v. France*, Legal Summary, No. 36109/03 (ECHR 2008).
8 *Kaptan v. Switzerland*, No. 55641/00 (ECtHR 2001).
9 *Hizb ut-Tahrir and Others v. Germany*, No. 31098/08 (ECtHR 2012).
10 *R. v. Iqbal*, 2014/01692 C5 (EWCA Crim. 2650, 2014).
11 *Mahajna v. Home Secretary*, UKUT B1 (IAC) (U.K. Upper Tribunal, Immigration and Asylum Chamber 2012).
12 *Ibid.*
13 *Hizb ut-Tahrir and Others v. Germany* (ECtHR)

explicitly praises past attacks.[14] Justification is considerably less likely for content that only implicitly justifies or approves of violence,[15] such as statements which joke about or ambiguously refer to attacks, without defending them as justified.[16] In these borderline cases, some courts weigh contextual factors more heavily. Others will refuse to weigh contextual factors when the content does not clearly communicate approval of violence. Even omissions (such as failures to condemn attacks by others) have sometimes been considered to justify restriction,[17,18] although this represents the exception rather than the general rule.[19]

As a result, the following functions for a post did not justify restriction in any of the 29 cases identified: sharing a VEO's non-violent political goals or supporting its underlying ideology[20,21,22] or criticizing government actions or policies.[23,24] Courts emphasize that criticism of the government on matters of public interest is entitled to particularly strong protection as free expression, and that the restraint governments must show in criminalizing speech makes imprisonment a wholly disproportionate response to criticism of public authorities.[25,26,27]

4. Any Violence Resulting from the Post

A post is more likely to justify restriction if violence appeared to result from the post,[28] and less likely if no violence resulted.[29,30] However, no court required that actual violence result from a post in order to deem a post incitement and justify restriction.[31]

14 *Case of Jose Miguel Arenas*, No. 79/2018, (Spanish Supreme Court 2018).
15 *The State v. Cassandra Vera*, STC 493/2018, (Spanish Supreme Court 2018).
16 E.g., The State v. Cassandra Vera, STC 493/2018, (Spanish Supreme Court 2018).
17 *Herri Batasuna and Batasuna v. Spain*, Nos. 25803/04 and 25817/04 (ECtHR 2009).
18 *Leroy v. France*, Legal Summary, No. 36109/03 (ECHR 2008).
19 *Lehideux and Isorni v. France*, No. 24662/94 (ECHR Grand Chamber, Sept. 23, 1998).
20 *The Case of the Academics for Peace*, No. 2018/17635 (Turkish Constitutional Court, Jul. 26, 2019).
21 *The Case of Ayse Celikel*, App. No. 2017/36722 (Turkish Constitutional Court, May 9, 2019).
22 *Keun-Tae Kim v. Republic of Korea*, No. 574/1994 (UN Human Rights Committee, Nov. 3, 1998).
23 *Fatih Tas v. Turkey* (No. 3), No. 45281/08 (ECtHR Grand Chamber, Apr. 24, 2018).
24 *Faruk Temel v. Turkey*, No. 45281/08 (ECtHR Second Section, Apr. 24, 2018).
25 *Ibid.*
26 *The Case of the Academics for Peace*, No. 2018/17635 (Turkish Constitutional Court, Jul. 26, 2019)
27 *The Case of Ayse Celikel*, App. No. 2017/36722 (Turkish Constitutional Court, May 9, 2019) [in Turkish].
28 *Keun-Tae Kim v. Republic of Korea*, No. 574/1994 (UN Human Rights Committee, Nov. 3, 1998).
29 *Mahajna v. Home Secretary,* UKUT B1 (IAC) (U.K. Upper Tribunal, Immigration and Asylum Chamber 2012).
30 *Vanjai v. Hungary*, No. 33629/06 (ECtHR, July 8, 2008).
31 *Case of Jose Miguel Arenas*, No. 79/2018 (Spanish Supreme Court, Feb. 15, 2018).

As correlation does not equal causation, courts did not necessarily assume that violence following a post proved that the post constituted incitement and therefore justified restriction. For instance, when violent protests broke out after a party official distributed pamphlets, the UN HRC held his conviction for distributing the pamphlets nevertheless violated free expression because the pamphlets advocated only peaceful actions.[32] In another instance, a court dismissed an allegation that a speech had led to a nearby riot, noting that a video of the speech itself showed no violence.[33]

5. Timing of the Post

A post is more likely to justify restriction if it is posted very soon after a VEO attack,[34] or during a very sensitive political situation, such as ongoing violent unrest.[35, 36] It is less likely to justify restriction if it apparently endorses an attack or crime perpetrated decades ago.[37,38]

6. Medium & Reach of the Post

Courts sometimes consider the medium of the content: whether support is being expressed via a newspaper, blogging website, social media.[39] The more public the platform, the more likely the speech is to justify restriction, based on the ability of the posts to reach mass audiences.[40] Editorials or articles, deliberately written and selected for publication,[41] might be more likely to support restriction than hasty spur-of-the-moment speech, such as a call-in comment to a broadcast.[42]

[32] *Keun-Tae Kim v. Republic of Korea*, No. 574/1994 (UN Human Rights Committee, Nov. 3, 1998)
[33] *Mahajna v. Home Secretary*, UKUT B1 (IAC) (U.K. Upper Tribunal, Immigration and Asylum Chamber 2012).
[34] *Leroy v. France*, Legal Summary, No. 36109/03 (ECtHR, Feb. 2, 2008).
[35] *The State v. Cassandra Vera*, STC 493/2018 (Spanish Supreme Court, Feb. 26, 2018).
[36] *Zana v. Turkey*, No. 40984/07 (ECHR Grand Chamber, Nov. 25, 1997).
[37] *Lehideux and Isorni v. France*, No. 24662/94 (ECtHR Grand Chamber, Sept. 23, 1998).
[38] *The State v. Cassandra Vera*, STC 493/2018 (Spanish Supreme Court, Feb. 26, 2018).
[39] *Smajić v. Bosnia and Herzegovina*, No. 48657/2016, (ECtHR 2018).
[40] *The State v. Cassandra Vera*, STC 493/2018, (Spanish Supreme Court, Feb. 26, 2018).
[41] *Fatih Tas v. Turkey* (No. 3), No. 45281/08 (ECtHR Grand Chamber, Apr. 24, 2018).
[42] *The Case of Ayse Celikel*, App. No. 2017/36722 (Turkish Constitutional Court, May 9, 2019) [in Turkish].

7. Speaker's Role and Personal History

Courts often consider a speaker's role, particularly in terms of his or her ability to exert authority or influence, e.g. as a leader of a political party,[43] former office-holder,[44] leader of an extremist group, or editor,[45] versus that of an academic,[46] private citizen,[47] or satirist.[48] Speech from a speaker with great prominence or a large number of followers is more likely to justify restriction.[49, 50]

Courts may also consider a speaker's prior behavior or personal history and circumstances when determining whether to ascribe knowledge or intent to incite violence to a piece of speech.[51, 52, 53]

8. Proportionality of Sentencing

Sentences of two years or longer in prison for ideological support offenses were almost uniformly held to violate freedom of expression,[54] especially when the speakers did not explicitly incite violence.[55] Only two courts upheld such prison sentences; a U.K. case involving encouragement that people travel to join ISIS,[56] and a Spanish case involved explicit justifications of violence and explicit praise of past violence.[57]

Even shorter prison or jail sentences for speech supporting VEOs were usually considered to violate freedom of expression.[58] Out of twenty-eight total cases, eleven involved sentences of

[43] *Keun-Tae Kim v. Republic of Korea*, No. 574/1994 (UN Human Rights Committee, Nov. 3, 1998).
[44] *Zana v. Turkey*, No. 40984/07 (ECtHR Grand Chamber, Nov. 25, 1997).
[45] *Fatih Tas v. Turkey* (No. 3), No. 45281/08 (ECHR Grand Chamber, Apr. 24, 2018)
[46] *The Case of the Academics for Peace*, No. 2018/17635 (Turkish Constitutional Court, Jul. 26, 2019)
[47] *Fatullayev v. Azerbaijan,* No. 40984/07, (ECtHR 2010).
[48] *Tribunal de grande instance de Paris (jugement correctionnel du 18 mars 2015)*, (High Court of Paris, Mar. 18, 2015)
[49] *Ibid.*
[50] *Zana v. Turkey*, No. 40984/07 (ECHR Grand Chamber, Nov. 25, 1997).
[51] *The Case of Dmitry Semenov*, 22-2559/2015, (Russian Appellate Ct. 2015).
[52] *Mahajna v. Home Secretary*, UKUT B1 (IAC) (U.K. Upper Tribunal, Immigration and Asylum Chamber 2012).
[53] *The State v. Cassandra Vera*, STC 493/2018, (Spanish Supreme Court, Feb. 26, 2018)
[54] E.g., *Stomakhin v. Russia*, No. 52273/07, (ECtHR 2017).
[55] E.g., *Fatullayev v. Azerbaijan*, No. 40984/07, (ECtHR 2010).
[56] *Choudary & Anor v. Regina*, England and Wales Court of Appeal (Criminal Division) (Mar 22, 2016).
[57] *Case of Jose Miguel Arenas*, No. 79/2018 (Spanish Supreme Court, Feb. 15, 2018).
[58] *The State v. Cassandra Vera,* STC 493/2018 (Spanish Supreme Court, Feb. 26, 2018).

imprisonment. Only three upheld these sentences as consistent with free expression,[59,60,61] The remaining eight held the prison sentences violated free expression.[62]

Two cases involved suspended prison sentences. One found the sentence proportionate for ideological support that fell short of incitement to violence,[63] while the other did not.[64]

The remaining cases involved civil penalties or criminal fines, which often did not violate free expression. Eight held the restrictions justified,[65] compared to six that found violations.[66]

B. Spreading False Information Online: Global Legislative Trends

1. International Jurisprudence on Criminalizing the Spread of False Information

In four key areas, international human rights law has established frameworks and guidance on whether criminalizing the spread of false information is permitted: (i) incitement to violence or hate speech, (ii) fraud or false advertising, (iii) defamation, and (iv) 'memory' laws concerning false information about historical occurrences. International human rights law jurisprudence clearly permits the criminalization of false information that rises to the level of fraud, incitement to violence, or hate speech, but generally prohibits the criminalization of mere defamation.

Beyond these four areas, authoritative opinions in international and regional human rights law have declared that other prohibitions of misinformation, to be legitimate, must pass the three-part test. Courts have often found that specific misinformation prohibitions fail the three-part test because they are too vague to be properly 'prescribed by law' and too broad or severe to be

59 *Choudary & Anor v. Regina* (England and Wales Court of Appeal, Criminal Division, Mar 22, 2016).
60 *Case of Jose Miguel Arenas*, No. 79/2018 (Spanish Supreme Court, Feb. 15, 2018).
61 *Zana v. Turkey*, No. 40984/07 (ECtHR Grand Chamber, Nov. 25, 1997).
62 E.g. *Fatullayev v. Azerbaijan*, No. 40984/07, (ECtHR 2010).
63 *Tribunal de grande instance de Paris (jugement correctionnel du 18 mars 2015)*, (High Court of Paris, Mar. 18, 2015).
64 *Smajić v. Bosnia and Herzegovina*, No. 48657/2016 (ECtHR 2018).
65 E.g. *Case of Saygili and Falakaoglu v. Turkey* (No. 2), No. 38991/02 (ECtHR Grand Chamber, Feb. 27, 2009); *R. v. Iqbal*, 2014/01692 C5 (EWCA Crim. 2650, 2014); *Leroy v. France, Legal Summary*, No. 36109/03 (ECHR, Feb. 2, 2008).
66 E.g. *Granier et al. v. Venezeula*, Report on Merits, Report No. 112/12, Case No. 12.828 (Inter-Am. Comm'n H.R., Nov. 9, 2012); *Vanjai v. Hungary*, No. 33629/06 (ECtHR, July 8, 2008).

necessary. Mere falsity of information, removed from its likelihood of causing harm, does not seem to justify criminalizing the spread of that information.

2. Global Trends in State Reponses to False Information

This study assessed the legislative responses combatting the spread of false information across 194 countries, organized according to ten geographic and political regions: Africa, Asia, the Caribbean, Central and South America, the Commonwealth of Independent States, Europe (non-EU), Europe (EU), the Middle East, North America, and Oceania. Notably, the below quantitative characterizations of this research do not differentiate between actions that are being processed, passed, implemented, and repealed by governments. However, proposed legislation was only included as an 'action taken or considered' when the study team could not verify whether the legislation had passed or not.

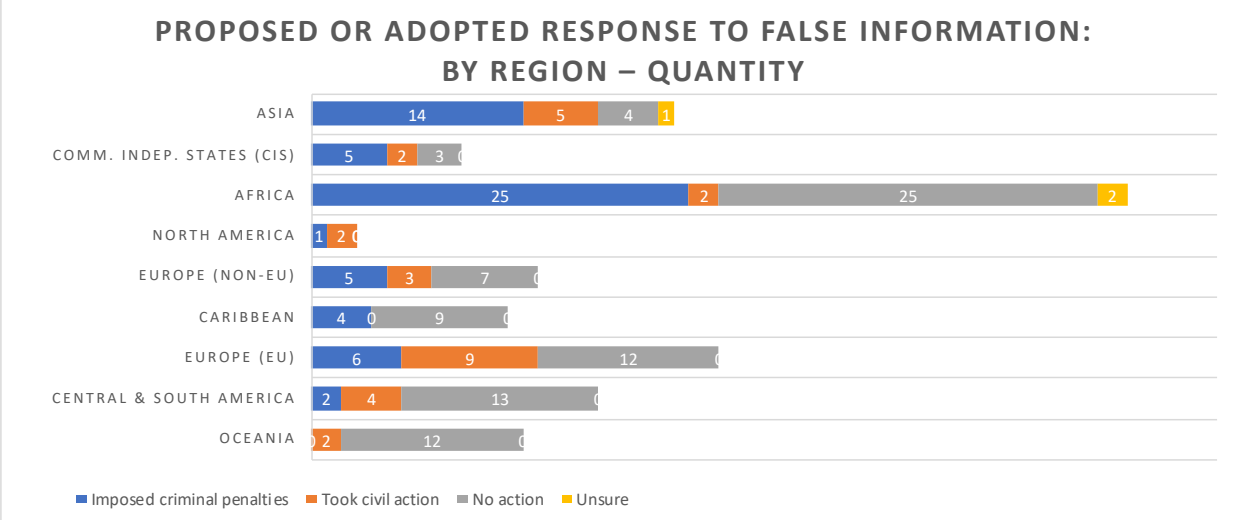
States respond to the spread of false information in a variety of ways. The term of "false information" here encompasses both misinformation (the spread of *unintentionally* false information) and disinformation (the spread of *intentionally* false information). States' legislative efforts to combat the spread of false information included imposing criminal penalties including fines and imprisonment on individuals who shared false information, imposing civil liability on those same individuals, imposing corporate liability on intermediaries like social media platforms who fail to remove false information, and funding awareness campaigns or task forces charged with countering the spread of false information.

Globally, out of the 192 surveyed, 101 countries have passed or considered legislative responses to the spread of false information. Taking all regions together, as determined by statistical analysis, it was common for at least some but not all countries in a region to consider taking some kind of legislative action against false information. A statistically typical region fell in a range of 27-77% of countries having taken or proposed some kind of action against false information.

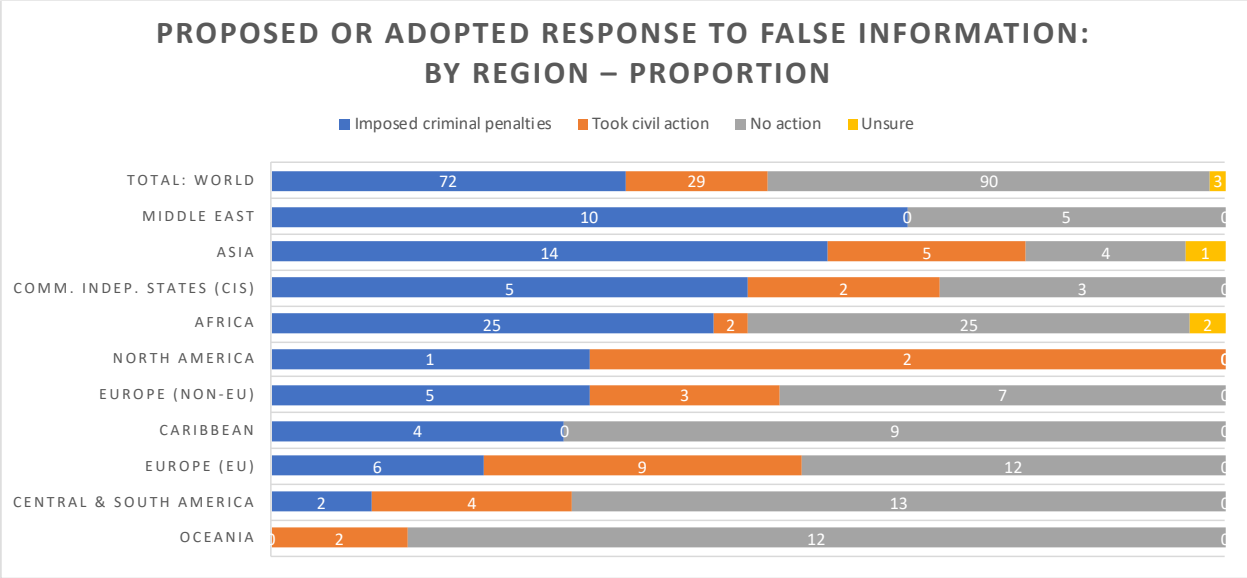
Approximately one-third of countries have criminalized or entertained legislative proposals criminalizing this activity. The majority have not. Of the 101 countries considering or taking action on this issue, 72 have passed or contemplated criminalizing the spread of false

information. Therefore, if a country did take legislative action against false information, it was likely to be criminal in nature; based on standard deviation from the mean, a statistically typical region fell in a range of 38-100% of countries which had considered or taken any type of action against the spread of false information to have included criminal penalties in that response.

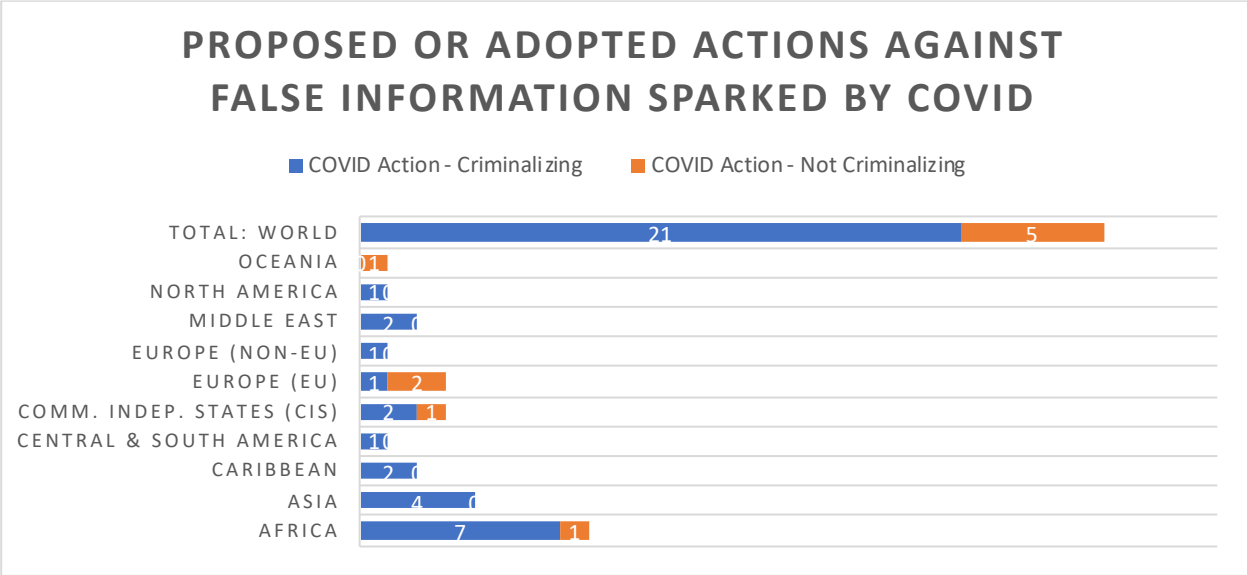
In most regions, a minority of countries had passed or considered laws criminalizing the spread of false information. The countries that have taken or considered criminalization measures are mostly in Asia and Africa, but these are also the regions with the highest numbers of countries. A majority of the states in the Middle East and the Commonwealth of Independent States have also considered or taken action to criminalize false information. In all other regions, criminalization is a minority response.



Worldwide, countries imposing or considering imposing criminal liability on individuals are in the minority. They represent 37% of all 192 countries surveyed. The only two regions in which a majority of countries have legislatively considered criminalizing the spread of false information were the Middle East and Asia. In two other regions, the proportion of criminalizing and non-criminalizing countries was roughly equal: Africa, slightly under half had proposed or passed criminal prohibitions, and in the Commonwealth of Independent States, half had. On the other end of the spectrum were five regions. In North America, the Caribbean, Central and South America, and Europe, countries contemplating criminal responses to false information were in the decided minority. In Oceania, no country had done so.



Notably, the COVID-19 pandemic has sparked about a quarter of all actions proposed or taken against false information, and the majority of those pandemic-related actions have included criminal penalties. 24 countries took or proposed their actions in response to COVID, and 21 of these proposals included criminalization. In other words, 24% of all the actions considered or taken against misinformation were in direct response to the COVID-19 pandemic, representing, and 88% of these COVID-19-prompted actions including the criminalization of false information. The number of countries who engaged in these actions is not large enough to support robust analysis of how these trends varied by region. As of today, there is no telling how long these laws will be applied and if these laws criminalizing misinformation will cease to be applied once the pandemic is over.



The phenomenon of state efforts to combat the online spread of false information is a new and developing area. As the response to the COVID-19 pandemic shows, state practice may evolve quickly in response to new conditions, so this snapshot of global practice should not be relied upon without considering the potential for state practice to shift.

C. Defamation Online: Jurisprudential and National Trends

The chapter on defamation online identifies three key trends in international legal jurisprudence on defamation.

1. Facts versus Opinion: To Be Defamatory, Statements Should be Factual rather than Opinion Based.

Since opinions are assertions whose truth cannot be verified, a number of courts have ruled that criminalizing a statement of opinion as defamation violates freedom of expression. The ECtHR established this doctrine in *Lingens v. Austria,* emphasizing the need for "a careful distinction . . . between facts and value judgements," as facts can be proved but value judgements cannot.[67]

[67] *Lingens v. Austria*, Application No. 9815/82 (ECtHR 1986).

2. Criminalization of Defamation: Despite State Practice to the contrary, International Courts Consider Criminalizing Defamation Too Severe a Form of Restriction to Comply with Free Expression.

Human rights bodies and courts have consistently asserted that criminalization of defamation is a disproportionate restriction that violates free expression. The ECtHR has repeatedly found applications of criminal defamation laws to disproportionately restrict free expression; the court accepts criminal penalties rather than civil restrictions for defamation only in extraordinary circumstances in which the speech at issue impairs other fundamental rights, such as when hate speech or incitement to violence are at play.[68] Similarly, the African Court of Human and Peoples' Rights, in 2014, ruled that criminal charges for libel are disproportionate responses that therefore violate freedom of expression, unless exceptional circumstances—defined as speech that constitutes incitement to violence or hate speech—are present.[69] The Inter-American Court of Human Rights (IACtHR) has also taken a strong stance against criminal defamation, especially in the context of journalists, and consistently adopts "precautionary measures" calling on member states to take immediate corrective action to aid journalists facing imprisonment.[70]

While many states have criminal defamation laws, state practice may indicate an emergent trend towards decriminalizing defamation. Since 2000, over 30 states have taken steps towards doing away with criminalizing defamation, and over 10 countries in Central and South America have abolished the imposition of prison sentences for defamation.[71]

Practice varies somewhat by region. In the Asia Pacific region, criminal defamation laws are widespread. Many countries in Europe also maintain criminal defamation laws, particularly in Southern Europe and Central Europe (such as Greece, Italy, Portugal, Turkey, Hungary, and

[68] Merita Kettunen, *Legitimizing European Criminal Law: Justification and Restrictions*, Springer Nature, Nov. 8, 2019.
[69] *Konaté v. Burkina Faso*, App. No. 004/2013 (Afr. Court Hum. Peoples Rights, Dec. 5, 2014). Available at: http://www.ijrcenter.org/wp-content/uploads/2015/02/Konate-Decision-English.pdf
[70] Alexandra Ellerbeck, "Inter-American Human Rights System, campaigns against defamation laws keep journalists from jail in Americas," Committee to Project Journalists, Dec. 15, 2015. Available at: https://cpj.org/2015/12/journalists-jail-inter-american-human-rights-defamation/
[71] Amicus Curiae Brief of Intl. Human Rights Clinic at Yale Law School, *In re Emilio Palacio Urrutia et al*, No. P-143611. Available at: https://law.yale.edu/sites/default/files/area/center/schell/commission_no._p-143611_amicus_lowenstein_clinic_english.pdf

Azerbaijan. Some countries in the Americas criminalize defamation, but generally countries in this region have been moving away from criminalizing defamation. A number of countries in

3. Criticism of Public Officials: Free Expression Requires States to Show Particular Restraint in Restricting Criticism of Public Figures and Heads of State.

International human rights bodies and courts agree that states must show greater restraint when contemplating a restriction on speech that criticizes public figures. For instance, in *Lingens* v. *Austria,* the ECtHR held that "[t]he limits of acceptable criticism are . . . wider as regards a politician as such than as regards a private individual;" since a politician knowingly opens up "his every word and deed" to close public scrutiny.[72] The ECtHR has also applied this doctrine towards inflammatory insults. In a similar vein, the IACtHR has ruled that public officials should be willing to undergo greater scrutiny than private individuals.[73] Some national courts have also adopted this position, including courts in Latin America and South America.[74]

Criminal defamation laws specifically targeting criticism of heads of state or other public officials are most common in Asia and—until recently—South and Central America. In Asia, *lèse majésté* laws criminalizing defamation or insult to heads of state are particularly widespread. Many *desacato* laws—defamation laws criminalizing speech that insults, threats, or injures public officials—were passed in Central and South America, but recently countries in this region have been moving away from them. In Central and South America, 9 countries have repealed their desacato laws.

4. Defamation of Religion: Criminal Blasphemy Laws May Comply with Free Expression Only When the Speech in Question Incites Discrimination, Hostility, or Violence.

International human rights jurisprudence and declarations suggest that criminal blasphemy laws generally violate rights to free expression, unless the speech restricted under them advocates religious hatred and incites discrimination, hostility, or violence. For instance, the UN Human

[72] *Lingens v. Austria*, Application No. 9815/82 (ECtHR 1986).
[73] *Case of Herrera-Ulloa v. Costa Rica* (IACHR, July 2, 2004).
[74] *See*, e.g., *Amparo Directo en Revisión*, 3123/2013 (Primera Sala de la Suprema Corte de Justicia de la Nación de México (SCJN), Feb. 7, 2014.

Rights Committee (HRC), monitoring body for the ICCPR, declared that "[p]rohibitions of displays of lack of respect for a religion or other belief system, including blasphemy laws," generally violate the ICCPR, with exceptions for advocacy of religious hatred, discrimination, and violence. In at least two cases, the ECtHR has held applications of criminal blasphemy laws to comply with freedom of expression, declaring that "expressions that seek to spread, incite or justify hatred based on intolerance, including religious intolerance" do not enjoy the protection of free expression under ECHR Article 10.

International human rights bodies such as the UN Human Rights Committee warn that using blasphemy laws to discriminate between religions or in favor of religion, or otherwise "to prevent or punish criticism of religious leaders or commentary on religious doctrine and tenets of faith," would violate free expression. A joint declaration from the OSCE, the UN, and the Inter-American and African human rights systems strikes similar notes, finding the concept of "defamation of religions" to be in conflict with international standards, reasoning that religions do not have their reputations in the way that individuals do.

State practice on blasphemy laws varies. Across the 194 countries in the world many do not have criminal blasphemy laws, and at least 4 countries have repealed such laws in the last 6 years. However, at least 46 countries have criminal blasphemy laws authorizing sentences of imprisonment.[75] These countries are distributed in across a number of regions, but notably do include almost no countries in North or South America or Oceania. Regions with larger numbers of countries who have criminal blasphemy laws include the Middle East and North Africa (11 countries), Sub-Saharan Africa (8 countries), Europe, (16 countries), Asia (7 countries), and the Commonwealth of Independent States (2). Additionally, 8 countries—mainly in the Middle East—authorize death sentences for engaging in criminal blasphemy.

[75] Algeria, Andorra, Austria, Bangladesh, Canada, Cyprus, Denmark, Finland, Egypt, El Salvador, Ethiopia, Gambia, Germany, Greece, Guyana, Ireland, India, Indonesia, Iraq, Israel, Italy, Jordan, Kazakhstan, Kuwait, Lebanon, Liechtenstein, Malaysia, Morocco, Oman, Pakistan, Poland, Portugal, the Russian Federation, Rwanda, San Marino, Spain, Sudan, Suriname, Switzerland, Tanzania, Thailand, Tunisia, Turkey, the United Kingdom (N. Ireland and Scotland only), Vatican City, and Western Sahara.

D. Cyberharassment and Cyberbullying: Jurisprudential and National Trends

1. Few Court Cases Found Applications of Cyberharassment Laws to Violate Free Expression

Research unearthed only three cases in which courts held that prosecutions under cyberharassment laws violated the right to free expression. In all three, the speech for which defendants were prosecuted was politically charged, and in two, the speech criticized public officials. One was decided under international principles by the OHCR's Working Group on Arbitrary Detention, in *Nyanzi v. Uganda*,[76] while the other two were decided by national courts under constitutional principles: the *Elliott* case in Canada[77] and the *Singhal* case in India.[78]

2. Many Forms of Cyberharassment Are at Least Lightly Criminalized

Cyberharassment activities tend to be at least lightly criminalized, i.e. eligible for statutory penalties of two years in prison or less. This includes cyberstalking, sexual harassment, and doxing. Cyberstalking generally entails repeated or persistent threatening conduct online with harmful effects for another. Sexual harassment often manifests online via distribution of sexual images without consent, often termed "revenge porn."

Many national laws also restrict other types of online conduct resulting in mental, psychological, physical, or economic harm. Countries are most likely to criminalize conduct that facilitates violence or physical damage, which sometimes falls under incitement to violence. However, countries also impose criminal sanctions for other harmful online behaviors that result in fear, alarm, annoyance, or distress. Many require that the perpetrator possess an intent to cause harm or knowledge that their behavior would cause harm. The publication of identity information

[76] *Opinion No. 57/2017 concerning Stella Nyanzi* (Human Rights Council Working Group on Arbitrary Detention 2017).
[77] *R. v. Elliott*, 2016 ONCJ 35 (Ontario Ct. of Justice, 2016). Available at: https://www.canlii.org/en/on/oncj/doc/2016/2016oncj35/2016oncj35.html
[78] *Shreya Singhal v. U.O.I*, Writ Petition No. 167 of 2012 (Sup. Ct. of India 2015). Available at: https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2015/06/Shreya_Singhal_vs_U.O.I_on_24_March_2015.pdf

that can cause similar harm (i.e. doxing) is a particular subset of this conduct that has been criminalized in multiple countries.

3. Cyberbullying by Minors Is Generally Not Criminalized

Numerous countries have passed or amended laws to address cyberbullying (online harassment by or of minors) in the context of their education systems. Cyberbullying undertaken by minors is generally not criminalized, except in some US states. Instead, most of the laws below require educational institutions to deal with the issue, while noting that if the issue is serious enough (usually, if it violates the criminal code) the police may get involved.

III. ARCHITECTURE OF INTERNATIONAL HUMAN RIGHTS LAW FOR ONLINE EXPRESSION

International human rights law (IHRL) protects the freedom of expression and the freedom of association, both offline and online. There are four primary sources of international law: conventions and treaties expressly agreed to by states, customary law as established by state practice, general principles of law recognized by states, and—as a subsidiary means of interpreting these rules of law—judicial decisions and the writings of experts in international law.[79] Conventions and treaties consist of both universal and regional mechanisms.

These sources of international law all establish that freedom of expression and association are core international human rights. However, they also clarify that these freedoms are subject to limitation under certain circumstances: when the restrictions are clearly stated in national law, necessary to protect national security, public order, public health, and public morality, or the rights and reputations of others, and proportionate to those aims.

The first section highlights the international human rights treaties and other mechanisms that are most relevant for freedom of expression and association. International mechanisms include the Universal Declaration of Human Rights (UDHR) as well as two treaties: the International Covenant on Civil and Political Rights (ICCPR) and the International Convention on the Elimination of All Forms of Racial Discrimination (ICERD). The key regional treaties are the European Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR), the American Convention on Human Rights (ACHR), and the African Charter on Human and Peoples' Rights (ACHPR). Regional instruments that are not yet binding include the and Arab Charter on Human Rights.

The second section presents key interpretive guidance of these treaties as issued by experts in international law. Of particular note, the Rabat Plan of Action articulates a six-factor test to assess whether an expression constitutes hate speech that rises to the level of inciting violence, including the (1) context of the statement, (2) position/status of the speaker, (3) intent of the

[79] Statute of the International Court of Justice, Article 38(1).

statement, (4) content of the statement, (5) extent/spread of the speech act, and (6) the likelihood of incitement.[80]

The third section articulates principles of jurisprudence derived from international and regional courts' interpretations of these treaties on freedom of expression and the permissibility of restricting it. These common principles include the applicability of free expression to the digital sphere and a three-part test that restrictions on free expression must pass in order to comply with international human rights law. They must be (i) prescribed by law, (ii) pursue legitimate aims, and (iii) be necessary and proportionate.

A. Relevant Treaties and Instruments of International Human Rights Law

1. International Mechanisms

Universal Declaration of Human Rights

The Universal Declaration of Human Rights (UDHR) was the first international human rights mechanism to articulate the fundamental rights now protected under international law.[81] The declaration is not a treaty, so its initial adoption did not itself create binding international law. However, many of its provisions have since been incorporated into customary international law, and therefore are now binding on all states.[82] The UDHR articulates freedom of expression and association as well as providing for how they may be restricted.

Enshrined in the UDHR is Article 19, which ensures the freedom of opinion and expression as well as the ability to seek out new information. Similarly, Article 20 protects the rights to

[80] "Rabat Plan of Action," *Annual Report of the United Nations High Commissioner for Human Rights*, A/HRC/22/17/Add.4, Jan. 11, 2013. Available at: https://www.ohchr.org/Documents/Issues/Opinion/SeminarRabat/Rabat_draft_outcome.pdf

[81] "70 Years of Impact: Insights on the Universal Declaration of Human Rights." unfoundation.org, Dec. 7, 2018. https://unfoundation.org/blog/post/70-years-of-impact-insights-on-the-universal-declaration-of-human-rights/. Kitsuron Sangsuvan, Balancing Freedom of Speech on the Internet under International Law, 39 N.C. J. Int'l L. & Com. Reg. 701 (2013). http://scholarship.law.unc.edu/ncilj/vol39/iss3/2.

[82] Hurst Hannum, "The Status of the UDHR in National and International Law," Ga. J. Int'l & Comp. L., Vol. 25: 287.

peaceful assembly and association.[83] In conjunction with Article 18, which ensures freedom to thought and religion, these "primary" articles form the bedrock for later human rights treaties.[84]

The UNHR contemplates, in Article 29, potential limitations of the rights and freedoms it articulates. This provision first provides that any limitations on the aforementioned rights and freedoms must be "determined by law," and have solely the purpose of "securing due recognition and respect for the rights and freedoms of others and of meeting the just requirements of morality, public order and the general welfare in a democratic society." In other words, the UNHR includes the first two prongs of the three-part test generally used to analyze restrictions on free expression (prescription by law and pursuit of legitimate aims), but does not include the third prong of necessity and proportionality. Additionally, the UNHR declares that the freedoms it enshrines "may in no case be exercised contrary to the purposes and principles of the United Nations."

International Covenant on Civil and Political Rights

The International Covenant on Civil and Political Rights (ICCPR), which was ratified in 1966, enshrines protections to freedom of speech and association as well as limitations to those freedoms. As the ICCPR is one of the most wide-ranging and broadly accepted human rights treaties, these restrictions have served as the legal basis for restrictions of speech and association across the globe.[85]

The ICCPR establishes the same fundamental rights as the UDHR: Article 18 (1, 2) ensures freedom of thought and religion; Article 19 (1, 2) ensures the freedom of opinions and expression, respectively; and Article 21 and 22 (1) ensure freedom of assembly and association, respectively. Each of these articles establishes freedoms, but also provides for their potential restrictions. Furthermore, the ICCPR contains two articles that further limit freedom of expression; Article 20, which obligates states to outlaw war propaganda and incitement of discrimination, hostility, and violence, and Article 5, which contains an anti-harboring provision.

[83] UN General Assembly, *Universal Declaration of Human Rights*, 10 Dec. 1948, 217 A (III), available at: https://www.refworld.org/docid/3ae6b3712c.html.
[84] *Ibid*., Art. 18.
[85] Elizabeth Cassidy, "Restricting Rights: The Public Order and Public Morality Limitations on Free Speech and Religious Liberty in UN Human Rights Institutions," *The Review of Faith & International Affairs*, 13:1, 5-12, 2015, DOI: 10.1080/15570274.2015.1005913.

The protections of the ICCPR are powerful. Article 19 provides that "[e]veryone shall have the right to freedom of expression," and explicitly says that "this right shall include freedom to seek, receive and impart information and ideas of all kinds," and that this applies "regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of [ ] choice." In other words, freedom of expression applies online as well as offline, and protects "ideas of all kinds," not merely those agreeable to governments.

Nonetheless, these protections come with significant and meaningful restrictions. Article 19 (3) recognizes that freedom of expression is not absolute and may be restricted, but only when the restriction is provided for by law and necessary for respect of "the rights and reputations of others" or to protect national security, public order, or public health or morals. In a similar fashion, Article 18 (3) provides for restrictions of the freedoms of religion and thought, Article 21 recognizes restrictions on the freedom of assembly, and Article 22 (2) authorizes restrictions on freedom of association.[86]

Beyond the freedom-specific restrictions for freedom of expression, two articles in the ICCPR provide for broader limitations on free speech.

First, Article 20 affirmatively obligates states to prohibit certain categories of expression. This provision declares that the following types of speech "shall be prohibited by law:" the use of "propaganda for war" and "[a]ny advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence"[87] However, the ICCPR's monitoring body for the ICCPR, the Human Rights Committee, has commented that state restrictions on expression falling under Article 20 must still comply with Article 19's three-part test for restrictions on speech.[88]

Second, Article 5 prevents an individual from exploiting the rights guaranteed in the ICCPR to insulate his attempt to undermine the rights of others from civil or criminal consequence. Specifically, Article 15 declares: "Nothing in the present Covenant may be interpreted as implying

[86] *International Covenant on Civil and Political Rights*, UN General Assembly, Dec. 16, 1966, United Nations, Treaty Series, vol. 999, p. 171. Available at: https://www.refworld.org/docid/3ae6b3aa0.html
[87] *Ibid*., Art. 20.
[88] *General comment no. 34, Article 19, Freedoms of opinion and expression*, UN Human Rights Committee (HRC), Sept. 12, 2011. Available at: https://www.refworld.org/docid/4ed34b562.html

for any State, group or person any right to engage in any activity or perform any act aimed at the destruction of any of the rights and freedoms recognized herein or at their limitation to a greater extent than is provided for in the present Covenant.[89] In other words, if one person seeks to destroy the ICCPR-guaranteed rights of someone else, he cannot rely on Article 19's right to freedom of expression to protect him.

International Convention on the Elimination of All Forms of Racial Discrimination

The International Convention on the Elimination of All Forms of Racial Discrimination (ICERD) is particularly relevant to prohibitions of hateful speech and hateful conduct.[90] Article 4 of this treaty commits member states to legally prohibit "all dissemination of ideas based on racial superiority or hatred, incitement to racial discrimination, as well as all acts of violence or incitement to such acts against any race or group of persons of another colour or ethnic origin," as well as the financing or other forms of assistance to "racist activities." This provision goes on to obligate joining states to legally prohibit organizations which "promote and incite racial discrimination" as well as other organized and "propaganda activities" which do the same, and make membership in those organizations "an offence punishable by law."

Other Core Human Rights Treaties and Monitoring Bodies

There are five other core international human rights treaties and monitoring bodies that are important to mention: the International Covenant on Economic, Social and Cultural Rights (ICESCR) and its monitoring body; the Convention on the Elimination of All Forms of Discrimination against Women (ICEDAW) and its monitoring body; the Convention on the Rights of Persons with Disabilities (ICRPD) and its monitoring body; the Convention on the Rights of the Child (ICRC) and its monitoring body; and finally, the International Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families (ICMW) and its monitoring body.

89 ICCPR, Art. 5.
90 *International Convention on the Elimination of All Forms of Racial Discrimination*, Office of the United Nations High Commissioner for Human Rights, 1965. Available at: https://www.ohchr.org/en/professionalinterest/pages/cerd.aspx

While these treaties and their monitoring bodies do not explicitly engage with freedom of expression concerns with the same robustness as the ICCPR or other regional human rights charters, they do inherently intersect with freedom of expression concerns as each population these treaties aim to protect are also protected under the ICCPR's Article 19, and other regional frameworks for free expression. For example, Article 21 of the Convention on the Rights of Persons with Disabilities asserts that State parties should take all appropriate measures to "ensure that persons with disabilities can exercise the right to freedom of expression and opinion, including the freedom to seek, receive and impart information and ideas on an equal basis with others and through all forms of communication of their choice."[91]

These core treaties and their monitoring bodies reassert the application of the ICCPR's Article 19, and free expression and opinion protection more broadly, to all individuals, and highlights the need for State parties to protect at-risk or marginalized populations. As the Human Rights Committee's General Comment #34 notes, highlighting voices from marginalized groups is essential for the free formation of opinion and free expression: "As a means to protect the rights of media users, including members of ethnic and linguistic minorities, to receive a wide range of information and ideas, States parties should take particular care to encourage an independent and diverse media."[92]

2. Regional Mechanisms

While the nearly universally applicable UDHR and ICCPR serve as the most wide-ranging and broadly applicable international human rights mechanisms, regional conventions also play a significant role in ensuring the protection of human rights around the world. These treaties reflect bindings agreements between states in different regions of the world: the European Convention for the Protection of Human Rights and Fundamental Freedoms, the American Convention on Human Rights, and the African Charter on Human and Peoples' Rights. Additionally, regional

91 *Convention on the Rights of Persons with Disabilities,* Article 21, U.N. Department of Economic and Social Affairs, 2009. Available at: https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities/article-21-freedom-of-expression-and-opinion-and-access-to-information.html.
92 *General comment no. 34, Article 19, Freedoms of opinion and expression*, UN Human Rights Committee (HRC).

human rights instruments such as the Association of Southeast Asian Nations (ASEAN) Human Rights Declaration do not bind participating states directly, but may nevertheless exert moral force.

European Convention for the Protection of Human Rights and Fundamental Freedoms

The European Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR), ratified in 1950, has served as the foundational example for subsequent human rights conventions, such as the American Convention on Human Rights.[93]

In the ECHR, Article 10(1) ensures freedom of expression, with restrictions to this right provided for by Article 10(2). Restrictions must be "prescribed by law" and be "necessary in a democratic society" for the purposes of advancing a legitimate aim. These potential legitimate aims include the advancement of "national security, territorial integrity, public safety, … the prevention of disorder or crime, … the protection of health and morals," as well as "the protection of the reputation or rights of others, . . . preventing the disclosure of information received in confidence, or . . . maintaining the authority and impartiality of the judiciary."

Likewise, Article 11(1) ensures freedom of assembly and association, subject to restrictions articulated in Article 11(2) on the similar grounds of national security, prevention of crime, protection of health, etc. Article 11(2) lacks the above-mentioned restrictions unique to expression, such as the protection of others reputation via defamation or libel legislation.

Additionally, Article 17 of the ECHR provides an important caveat to the freedoms enshrined in Article 10 and 11. Akin to Article 5 of the ICCPR, Article 17 of the ECHR ensures that none of the rights protected under the ECHR can be used to destroy or limit another's use of those rights. For instance, Article 10 does not specifically authorize restrictions on the basis of hate speech, but the European Court of Human Rights has interpreted Article 17 as making hate speech ineligible for the protections Article 10.[94] As a result, speech that promotes or justifies acts amounting to hatred, violence, xenophobia and racial discrimination, anti-Semitism, Islamophobia, terrorism and war crimes, as well as negation and revision of clearly established

[93] *European Convention for the Protection of Human Rights and Fundamental Freedoms*, as amended by Protocols Nos. 11 and 14, Council of Europe, Nov. 4, 1950. Available at: https://www.refworld.org/docid/3ae6b3b04.html.
[94] *Norwood v. United Kingdom*, App No. 23131/03 (ECtHR, July 16, 2004); *Pavel Ivanov v. Russia*, App. No. 35222/04 (ECtHR, Dec.18, 1996.

historical facts with the same effect, may be considered "unprotected speech," deprived of protection under Article 10 because it violates Article 17.

American Convention on Human Rights

Another regional human rights treaty is the American Convention on Human Rights (ACHR), which had its last ratifying country in 1997.[95] The ACHR contains standard provisions protecting freedom of speech and freedom of association/assembly in Articles 13 (1) and 16 (1), as well as outlining the requirements for any restrictions on those rights. However, the ACHR lacks a more flexible provision in the style of ECHR Article 17 to implicitly exclude, from protected status, actions aimed to prevent the exercise of others' rights.[96]

Similarly to the ICCPR and ECHR, Articles 13 and 16 authorize limitations on freedoms of speech, association, and assembly, if those limitations are "prescribed by law" and "necessary in a democratic society" to ensure the same aims articulated in the ICCPR: "the protection of national security, public order, or public health and morals" as well as, for freedom of expression: "respect for the rights and reputations of others."[97] Notably, the ACHR also adds that these limitations may not operate via "prior censorship," other than as applied to regulating access by minors to public entertainment.

Article 13(5) also specifically outlaws "propaganda for war" and hate speech, defined as "any advocacy of national, racial, or religious hatred that constitute incitements to lawless violence or to any other similar action against any person or group of persons on any grounds including those of race, color, religion, language, or national origin."[98]

African Charter on Human and Peoples' Rights

[95] *American Convention on Human Rights*, Organization of American States (OAS), Nov. 22, 1969. Available at: https://www.refworld.org/docid/3ae6b36510.html
[96] *Ibid.*
[97] *Ibid.*
[98] *Ibid.*

Finally, the last widely recognized regional human rights mechanism is the African Charter on Human and Peoples' Rights (ACHPR), ratified in 1979.[99] Under Articles 9 and 10 of the ACHPR respectively, freedom of speech and freedom of association are protected.

In contrast to the ECHR, ACHR, and ICCPR, the ACHPR provides no explicit limitations to expression or association. Instead, Article 27 of the ACHPR emphasizes an individual's duties to "his family and society, the State and other legally recognized communities and the international community" and their obligation to exercise their rights "with due regard to the rights of others, collective security, morality and common interest."[100]

Beyond emphasizing the expectations for responsibly exercising one's rights, the ACHPR explicitly outlines individuals' duties in a variety of contexts. In particular, Article 28 outlines the obligation to protect others from discrimination, with Article 29 (3) accentuating the importance of protecting a state's security and Article 29 (6) state the necessity "to contribute to the promotion of the moral wellbeing of society." This collection of duties and obligations has been interpreted, similar to Article 17 of the ECHR, to restrict hate speech and similar actions that could infringe upon the rights of others. However, its lack of a specific clause explicitly outlawing such actions has led to lax enforcement compared to the ECHR.[101]

The Association of Southeast Asian Nations Human Rights Declaration

The Association of Southeast Asian Nations (ASEAN) Human Rights Declaration is not a legally binding treaty, but may nevertheless exert moral force on the states that issued it in 2012. The declaration affirms the right to free expression—but not association—and provides for broader exceptions than the above-mentioned human rights treaties.

Article 23 affirms the right to free expression, declaring that all people have "the right to freedom of opinion and expression," which includes the "freedom to hold opinions without interference" as well as "to seek, receive and impart information, whether orally, in writing or

[99] *African Charter on Human and Peoples' Rights ("Banjul Charter")*, Organization of African Unity (OAU), June 27, 1981. Available at: https://www.refworld.org/docid/3ae6b3630.html
[100] *Ibid.*
[101] B. Obinna Okere, "The Protection of Human Rights in Africa and the African Charter on Human and Peoples' Rights: A Comparative Analysis with the European and American Systems." *Human Rights Quarterly* 6, no. 2 (1984): 141-59. doi:10.2307/762240.

through any other medium of that person's choice." Article 22 affirms "the right to freedom of thought, conscience and religion." Correspondingly, Article 22 explicitly authorizes restrictions on free expression by declaring that "[a]ll forms of intolerance, discrimination and incitement of hatred based on religion and beliefs shall be eliminated." [102]

The ASEAN declaration provides for the restriction of these freedoms in a broad catch-all provision mandating that limitations be "determined by law solely for the purpose of securing due recognition for the human rights and fundamental freedoms of others," and for the purpose of meeting "the just requirements of national security, public order, public health, public safety, public morality," and "the general welfare of the peoples in a democratic society." The exceptions to free expression therefore authorized are broader than those in the ECHR and other human rights treaties in two ways: first, the declaration pens up the list of legitimate aims to include the "general welfare," and second, it does not include an explicit requirement of necessity.

Arab Charter on Human Rights

Adopted at the 2004 Arab Summit, the Arab Charter on Human Rights articulates freedoms of expression and association as well as providing for their restriction. This treaty will not enter into force until at least seven states complete ratification. [103]

The Arab Charter's provision on free expression, Article 32, substantially mirrors ICCPR's Article 19. The provision ensures the rights to "freedom of opinion and freedom of expression," as well as the "freedom to seek, receive and impart information by all means, regardless of frontiers." This article also provides for the restriction of these rights, but only those "necessary for the respect of the rights or reputation of others," or "for the protection of national security or of public order, health or morals." However, the requirements for restrictions do not include that the restrictions be provided for by law.

The Arab Charter's Article 24 similarly enshrines the right of "[e]very citizen" to engage in "political activity," peacefully assemble, and form and join associations with others. Restrictions

[102] *ASEAN Human Rights Declaration*, Association of Southeast Asian Nations, 2012. Available at: https://asean.org/asean-human-rights-declaration/
[103] Mohammed Amin Al-Midani, "Arab Charter on Human Rights 2004," *Boston Univ. Int'l L.J.*, 2006, Vol. 24:147. Available at: http://www.eods.eu/library/LAS_Arab%20Charter%20on%20Human%20Rights_2004_EN.pdf

are permitted under the similar circumstances as for freedom of expression. However, in contrast to the restrictions for freedom of expression, this provision does require that restrictions be "imposed in conformity with the law."

B. Relevant Guidance and Interpretations from International Bodies and Experts

Recent guidance documents from United Nations officials and outside international law experts address how to best protect human rights in the modern era. These include general comments issued by the Human Rights Committee, which monitors state implementation of the ICCPR; reports from the UN Human Rights Council's Special Rapporteur for Freedom of Expression; the Rabat Plan of Action, which was endorsed by the UN Office of the High Commissioner for Human Rights; and the Johannesburg Principles on National Security, Freedom of Expression and Access to Information adopted by a conference of human rights experts.

1. Human Rights Committee

The Human Rights Committee, a body of independent experts, monitors the implementation of the ICCPR by its State parties. In addition to holding sessions in Geneva and assessing State party reports on how the rights of the ICCPR is being implemented in their country, the Human Rights Committee issues interpretation of the content of the provisions within the ICCPR, known as general comments on thematic issues or its methods of work. General Comment #34 (replacing previous General Comment #10), offers interpretational guidance for Article 19, and generally how State parties can observe and protect freedom of opinion and expression.[104]

In the comment, the Human Rights Committee emphasizes that Article 19 extends to "internet-based modes of expression," and provides further clarity to the exceptions to free expression, which must be "provided by law," and necessary legitimate aims (e.g. protecting the "rights and reputations of others"). The comment also underscores that this interpretation extends to the digital: "Any restrictions on the operation of websites, blogs or any other internet-based, electronic or other such information dissemination system, including systems to support such

[104] *General comment no. 34, Article 19, Freedoms of opinion and expression*, UN HRC.

communication, such as internet service providers or search engines, are only permissible to the extent that they are compatible with paragraph 3 [of Article 19 of the ICCPR]."

Finally, the comment provides specific interpretation relevant to counter-terror restrictions by State parties, explaining that "such offences as 'encouragement of terrorism' and 'extremist activity' as well as offences of 'praising', 'glorifying', or 'justifying' terrorism, should be clearly defined to ensure that they do not lead to unnecessary or disproportionate interference with freedom of expression….The media plays a crucial role in informing the public about acts of terrorism and its capacity to operate should not be unduly restricted." This suggests that states should strictly to what is commonly referred to as the "three part test" (provided by law, necessary, and for legitimate aims) as described in Article 19.

2. United Nations (UN) Entities

Reports from UN officials suggest three principles relevant for determining whether criminally punishing an individual's social media post complies with international freedom of expression. First, the Secretary-General's report on protecting human rights while combatting terrorism recommends distinguishing between glorification and incitement: states should only criminally prosecute "direct incitement to terrorism," defined as "speech that directly encourages the commission of a crime, is intended to result in criminal action and is likely to result in criminal action."[105] Second, the Special Rapporteur on Free Expression considers audience size a salient consideration when considering how likely a post is to incite violence: "a statement released by an individual to a small and restricted group of Facebook users does not carry the same weight as a statement published on a mainstream website."[106] Third, the Rapporteur also instructs that artistic

[105] *The protection of human rights and fundamental freedoms while countering terrorism*, Report of the Secretary-General, UN General Assembly, A/63/337, Aug. 28, 2008. Available at: https://www.securitycouncilreport.org/atf/cf/%7B65BFCF9B-6D27-4E9C-8CD3-CF6E4FF96FF9%7D/Terrorism%20A%2063%20337.pdf

[106] *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, UN Human Rights Council, A/HRC/67/357, Sept. 7, 2012. Available at: https://undocs.org/pdf?symbol=en/A/67/357.

merit should be assessed in context, "given that art may be used to provoke strong feelings without the intention of inciting violence, discrimination or hostility."[107]

3. Rabat Plan of Action

The Rabat Plan of Action, issued by the UN Office of the High Commissioner for Human Rights in 2013, demonstrates modern methods for evaluating hate speech. In particular, the Rabat Plan of Action uses a six-part test for assessing whether an expression constitutes criminal hate speech that rises to the level of inciting violence, including the (1) context of the statement, (2) position/status of the speaker, (3) intent of the statement, (4) content of the statement, (5) extent/spread of the speech act, and (6) the likelihood of incitement.[108] This framework can help better ground the application of existing treaties like the ICCPR, ACHR, and ECHR to the speech of dangerous groups on social media platforms.

4. Johannesburg Principles on National Security, Freedom of Expression and Access to Information

The Johannesburg Principles on National Security, Freedom of Expression and Access to Information are a set of standards adopted by a conference of human rights experts in 1995.[109] While these Principles are not legally binding, nor have any countries made a formal or informal commitment to the Principles, they serve as another set of criteria similar to the Rabat Plan of Action that should guide decisions to restrict expression. Specifically, they offer guidance on what entails "national security" concerns and "necessary" action when restricting expression, placing the burden of proof on the government restricting expression.

First, the Johannesburg Principles interpret "necessity in a democratic society" as requiring that the following three conditions are met:

[107] *Ibid.*
[108] "Rabat Plan of Action," *Annual Report of the United Nations High Commissioner for Human Rights*, A/HRC/22/17/Add.4, Jan. 11, 2013. Available at: https://www.ohchr.org/Documents/Issues/Opinion/SeminarRabat/Rabat_draft_outcome.pdf
[109] The standards have been relied upon by judges, lawyers, and academics. Toby Mendel, The Johannesburg Principles: Overview and Implementation, Article 19, 7 Feb. 2003, available at: https://www.article19.org/data/files/pdfs/publications/jo-burg-principles-overview.pdf

1. "the expression or information at issue poses a serious threat to a legitimate national security interest;"
2. "the restriction imposed is the least restrictive means possible for protecting that interest; and"
3. "the restriction is compatible with democratic principles."

Second, the Principles suggest that a piece of expression may only be restricted "as a threat to national security" only if:

1. "the expression is intended to incite imminent violence;"
2. "it is likely to incite such violence; and"
3. "there is a direct and immediate connection between the expression and the likelihood or occurrence of such violence."[110]

C. Derived Jurisprudential Principles

A number of principles are common to the jurisprudence of almost all international regional courts when analyzing the permissibility of restrictions on free expression under international law. Many national courts also use these principles when analyzing alleged violations of national or constitutional rights to free expression.

The first common principle, which is universal enough to usually be tacitly assumed rather than explicitly articulated, is that the freedom of expression and association apply to online activities as well as offline activities. However, while some courts call for treating online and offline speech equally, others justify more severe restrictions on online speech. Still others urge greater caution when interfering with online speech than offline speech.

The second common principle involves the three-part test that courts apply when determining whether or not a restriction on speech is permissible.

Some courts, particularly the ECtHR, sometimes apply a threshold test to determine whether or not a piece of speech is eligible for protection by the right to free expression. In other

[110] *The Johannesburg Principles on National Security, Freedom of Expression and Access to Information*, Article 19, Oct. 1, 1995. Available at: https://www.refworld.org/docid/4653fa1f2.html

words, if a piece of speech aims to destroy the rights of others, the court may uphold the sanction without analyzing it under the three-part test explained below.

However, most courts apply the three-part test in every case without first applying this threshold qualification. The three requirements for a restriction on free expression or association under this test are: (1) prescription by law, (2) pursuit of a legitimate aim, and (3) necessity and proportionality. A criminal penalty may be considered disproportionate even when a civil penalty or restriction would not be. In general, at least for speech on matters of public interest, criminalization is viewed as a disproportionate form of restriction unless the speech constitutes incitement to hatred, discrimination, or violence.

1. Applicability of International Human Rights Law to Online Activities

International and regional human rights law establishes that the provisions on the freedom of expression apply regardless of the media of communication. However, there is some disagreement as to whether online speech should be treated exactly the same as offline speech; some entities call for equal treatment, others justify more severe restrictions for online speech, and still others urge greater caution before restricting online speech as compared to offline.

Some international human rights instruments explicitly affirm the universality of their applicability. The UDHR, ICCPR, ECHR, and ACHR state that the right applies "regardless of frontiers," and the American Declaration and Arab Charter articulate that it applies "by any medium whatsoever" and "through any medium."

The interpretation of equal treatment for offline and online speech has been embraced by many international bodies. The United Nations Human Rights Council (UNHRC) similarly declared that "the same rights that people have offline must also be protected online, in particular freedom of expression, which is applicable regardless of frontiers and through any media of one's choice."[111] The Human Rights Committee, the body of independent experts that monitors the implementation of the ICCPR, issued its interpretational guidance in General Comment #34, which

[111] Somini Sengupta, "U.N. Affirms Internet Freedom as a Basic Right," *The New York Times,* July 6, 2012. Available at: https://bits.blogs.nytimes.com/2012/07/06/so-the-united-nations-affirms-internet-freedom-as-a-basic-right-now-what

emphasized that Article 19 extends to "internet-based modes of expression."[112] The comment also underscores that the three-part test for restrictions on free expression (detailed below) extends to the digital sphere, i.e. to "[a]ny restrictions on the operation of websites, blogs or any other internet-based, electronic or other such information dissemination system, including systems to support such communication, such as internet service providers or search engines."[113]

The principle that human rights including the freedom of expression apply equally online and offline has also been affirmed by some international courts. For instance, in 2002, the Inter-American Court of Human Rights (IACtHR) stated that "the right to freedom of expression in the terms established by Article 13 of the American Convention equally protects both traditional media and the widespread expression via the Internet."[114]

Other international courts have found the unique aspects of the Internet justify greater restrictions on online free expression when compared to offline speech. For instance, in *Editorial Board of Pravoye Delo and Stekel,* the ECtHR concluded that the Internet can be a uniquely risky medium and thereby justify more severe restrictions. Specifically, the ECtHR found that Internet content and communications poses a higher "risk of harm . . . to the exercise and enjoyment of human rights and freedoms, particularly the respect for private life . . . than that posed by the press."[115] As a result, the ECtHR concluded that "the policies governing reproduction of material from the printed media and the Internet may differ," as policies for the Internet "have to be adjusted according to the technology's specific features" to adequately protect and promote rights and freedoms.[116]

Still other entities urge that states exercise even greater caution when enacting restrictions on online speech because such restrictions may have far-reaching knock-on effects. For instance, the Brazil Superior Court of Justice ruled against prior filtering of searches out of concern that the restriction of search results would harm all users and the internet overall, significantly outweighing

[112] *General comment no. 34, Article 19, Freedoms of opinion and expression*, UNHRC.
[113] *Ibid.*
[114] *Caso No. 12.367 - "La Nacion"* (IACHR 2002). Available at: http://www.corteidh.or.cr/docs/casos/herrera/demanda.PDF
[115] *Case of Editorial Board of Pravoye Delo and Shtekel v. Ukraine,* App. No. 33014/05 (ECtHR, May 5, 2011).
[116] *Ibid.*

the benefits a few users would gain.[117] Similarly, the OAS Special Rapporteur for Freedom of Expression argues that although freedom of expression enjoys the same protection in all media, it is important to carefully consider the potential impact of interfering with freedom of expression on the internet (e.g. comparing a decision's impact on the whole internet versus the impact on the victim).[118]

As a result, although some courts and international bodies apparently treat online speech differently, it is not clear whether these entities generally afford greater or lesser protection to online speech. Furthermore, it is not clear whether the entities that view online speech differently constitute a dissenting minority or an emerging norm deviating from the online equals offline approach suggested by international laws and treaties and other court rulings. This will likely be clarified as more rulings and legislation involving online speech proliferate.

2. Prohibitions on the Abuse of Rights: Eligibility for Protection

Sometimes, before applying the below three-part test to analyze the permissibility of a restriction on a piece of expression, courts interpreting the ECHR apply a threshold test under ECHR Article 17 to determine whether or not a piece of expression was eligible for ECHR Article 10 or Article 11 protection for free expression or association. The ICCPR has a similar provision, Article 5, which broadly restricts the use of rights protected in the Covenant to impede or limit the rights of others under the Covenant.

In the ECHR, this exemption from protection is contained in Article 17, "Prohibition of abuse of rights."[119] ECHR Article 17 declares that actions "aimed at the destruction of any of the rights and freedoms set forth" elsewhere in the ECHR are ineligible for the protections of the

[117] Edison Lanza, *National Case Law on the Freedom of Expression*, Office of the Special Rapporteur for Freedom of Expression of the Inter-American Commission on Human Rights, Organization of American States, Mar. 15, 2017. Available at: http://www.oas.org/en/iachr/expression/docs/publications/JURISPRUDENCIA_ENG.pdf

[118] Catalina Botero Marino, *Annual Report of the Inter-American Commission on Human Rights: Annual Report of the Office of the Special Rapporteur for Freedom of Expression*, Organization of American States, Dec. 31, 2013. Available at: http://www.oas.org/en/iachr/expression/docs/reports/annual/2014_04_22_%20IA_2013_ENG%20_FINALweb.pdf

[119] Full Text of Article 17: "Nothing in this Convention may be interpreted as implying for any State, group or person any right to engage in any activity or perform any act aimed at the destruction of any of the rights and freedoms set forth herein or at their limitation to a greater extent than is provided for in the Convention." ECHR.

ECHR. According to the ECtHR, the purpose of Article 17 is to prevent totalitarian organizations from exploiting the freedoms enshrined in the ECHR. Therefore, when a speaker directs a remark "against the Convention's underlying values," he forfeits eligibility for Article 10 protection of free expression, and courts need not apply the three-part test to see if the restriction is permissible.[120]

3. Three-Part Test: Prescription, Legitimate Aim, Necessity and Proportionality

Almost all courts articulate the same basic rule for analyzing a restriction on the freedom of expression or association online. This included international courts interpreting international treaties (including courts interpreting the ICCPR, the ECtHR interpreting ECHR, the IACtHR interpreting the ACHR, and ECOWAS interpreting the ACHPR), as well as domestic courts analyzing international or national constitutional principles of free expression. Guidelines about the interpretation of these conditions that are summarized here have been expanded upon by numerous courts, international bodies, and special rapporteurs.[121]

The basic three-part test is as follows. In order to justify a restriction on freedom of expression or freedom of association, the restriction must be (1) prescribed by law, (2) pursue a legitimate aim (such as national security, public safety, prevention of disorder, protection of the rights and freedoms of others), and be (3)(a) necessary in a democratic society as well as (3)(b) proportionate.

Prescription by Law

The first condition that any restriction must meet is legality; the restriction must be provided for by law. It must be established by formal legislation or its equivalent, adopted by regular legal processes (e.g. not based on traditional, religious, or other forms of customary law). The law must be accessible and precise, clearly distinguishing between what is lawful and unlawful such that citizens have notice of what is prohibited and what is permitted. Furthermore, it must be

120 *Hizb ut-Tahrir and Others v. Germany*, No. 31098/08, ¶ 72 (ECtHR 2012) (citing *Paksas v. Lithuania* [GC], no. 34932/04, § 45, §§ 87-88, 6 Jan. 2011).

121 Nicola Wenzel, "Opinion and Expression, Freedom of, International Protection," *Max Planck Encyclopedias of International Law,* Apr. 2014. Available at: https://opil.ouplaw.com/view/10.1093/law:epil/9780199231690/law-9780199231690-e855

applied by an independent body that can provide safeguards against its abuse, rather than being subject to unfettered executive discretion.

Legitimate Aim

The second condition is legitimacy; the restriction must pursue legitimate aims. The list of potential legitimate aims varies slightly from treaty to treaty, but generally echoes the list articulated in the ICCPR. ICCPR Article 19 and IACHR both identify the exact same list of legitimate aims: respecting "the rights and reputations of others" or protecting national security, public order, or public health or morals. Notably, the UNHCR has noted that states must abide by non-discrimination and universality principles while considering issues related to public morals by ensuring that they do not come from a "single tradition" of morality.[122] ECHR Article 10 has a longer but substantively similar list to that of the ICCPR, adding "preventing the disclosure of information received in confidence" and "maintaining the authority and impartiality of the judiciary," as well as expanding "national security" into "national security, territorial integrity or public safety." While the ACHPR has no clause explicitly providing for limitations to free expression or association, the Economic Community of West African States (ECOWAS Court) uses the same list of legitimate aims as in the ICCPR in its jurisprudence.[123]

Necessity and Proportionality

The third and final set of conditions are necessity and proportionality. The restriction must protect the legitimate interest of concern while imposing the lowest possible burden on the exercise of freedom of expression. In other words, a state must show a restriction was "necessary in a democratic society" (that there was a "pressing social need" for the restriction) and that the restriction effect was "proportionate to the legitimate aims pursued."[124] According to the UNHRC,

[122] Katie Bresner, "Understanding the Right to Freedom of Expression," *Journalists for Human Rights,* 2015. Available at: https://jhr.ca/wp-content/uploads/2019/10/Understanding-Freedom-of-Expression-Primer-ENG-web.pdf

[123] *Federation of African Journalists (FAJ) and others v. The Gambia*, (Community Court of Justice of the Economic Community of West African States, Apr. 2018). Available at: https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2016/04/FAJ-and-Others-v-The-Gambia-

[124] *The Observer and the Guardian v. United Kingdom* (ECtHR, 1991). Available at: https://www.ucpi.org.uk/wp-content/uploads/2018/03/The-Observer-and-The-Guardian-v-United-Kingdom-Application-No.-1358588-1992-14-E.H.R.R.-153.pdf

states must demonstrate the necessity and proportionality of the restriction by establishing a "direct and immediate connection between the expression and the threat."[125]

Assessment of proportionality generally involves holistically weighing the facts and circumstances of an offense against the severity of the sentence (or other form of interference) imposed by the state. Courts sharply distinguish between the permissibility of restrictions that merely involve civil consequences, including payment of damages to a plaintiff or loss of license, and those that impose criminal consequences, especially imprisonment, and find that governments must show restraint in criminalizing speech. Additionally, courts emphasize that criticism of the government on matters of public interest is entitled to particularly strong protection as free expression. Therefore, courts generally find that the restraint governments must show in criminalizing speech makes imprisonment a wholly disproportionate response to criticism of public authorities.[126,127,128]

Subject to two major caveats, the majority of courts seem to agree that speech—at least speech which is on matters of public interest—may only be criminalized if it constitutes incitement to violence, hostility, or discrimination.[129] The first caveat is that this limit on criminalization only applies to speech criminalized as pure speech, rather than conduct that happens to manifest via speech, such as threats, fraud, doxing, or stalking, etc. The second caveat is that states frequently seem to deviate from this norm, given the large number of states that have criminalized defamation.

[125] *General comment no. 34, Article 19, Freedoms of opinion and expression*, UN Human Rights Committee (HRC).
[126] *Faruk Temel v. Turkey*, No. 45281/08 (ECtHR Second Section, Apr. 24, 2018).
[127] *The Case of the Academics for Peace*, No. 2018/17635 (Turkish Constitutional Court, Jul. 26, 2019).
[128] *The Case of Ayse Celikel*, App. No. 2017/36722 (Turkish Constitutional Court, May 9, 2019) [in Turkish].
[129] *See, e.g., Case of Savva Terentyev v. Russia*, No. 10692/09 (ECtHR 2019) ("The imposition of a prison sentence for an offence in the area of a debate on an issue of legitimate public interest will be compatible with freedom of expression as guaranteed by Article 10 of the Convention only in exceptional circumstances, notably where other fundamental rights have been seriously impaired, as, for example, in the case of hate speech or incitement to violence."). *See also Kimel v. Argentina* (IACHR 2008); *Issa Konate v. The Republic of Burkina Faso*, No. 004/2013 (African Court on Human and People's Rights, Dec. 5, 2014). Available at: http://www.ijrcenter.org/wp-content/uploads/2015/02/Konate-Decision-English.pdf

IV. A NOTE ON THE RELEVANCE OF SOCIAL MEDIA PLATFORM POLICIES

As influential and omnipresent as social media platforms seem in the world's information environment, platforms are not actual governments, so their actions do not themselves constitute state practice which can indicate the formation of customary international law. However, platform policies often may reflect international norms, while exert powerful influence of their own on the formation of new international norms and state practice.

Since social media companies are at the forefront of confronting the increasingly pressing challenge of "regulating" online activities, each chapter includes a summary of how platforms' community guidelines and terms of service agreements approach these issues. Legal scholars Robert Chesney and Danielle K. Citron deem these platform policies "the single most important documents governing digital speech in today's world."[130]

These guidelines may be a harbinger of legal trends, as technology companies are often the first ones to encounter and adapt to the new reality of online speech. After all, these platforms have become what the UN Special Rapporteur on the promotion and protection of the rights to freedom of opinion and expression David Kaye refers to as "institutions of governance, complete with generalized rules and bureaucratic features of enforcement" whose decisions "influence public space, public conversation, democratic choice, access to information, and perception of the freedom of expression."[131]

The guidelines platforms employ may be used in the creation or interpretation of legislation, since they have become the norms that billions of people who use these platforms abide by. On the other hand, these guidelines suggest that alternatives to criminalization, such as takedown laws, may be emerging as the preferred approach to dealing with harms posed by hateful conduct, as many of these companies already take down a broad range of content.

These companies operate on a global scale, requiring them to comply with localized understandings of speech and expression. As a result, they are unable to fully shelter behind the

130 Danielle Citron and Robert Chesney, "Deep Fakes: A Looming Challenge for Privacy," *California Law Review* 107, no. 6 (2019): 1817, https://doi.org/10.15779/Z38RV0D15J.
131 David Kaye, *Speech Police: The Global Struggle to Govern the Internet* (New York: Columbia Global Reports, 2019).

speech laws of the countries in which they are headquartered. For instance, in the *Die Grünen* case, an Austrian court rejected Facebook's claim that it was governed only by laws in California or Ireland and compelled it to take down the post in issue worldwide.[132]

[132] *Die Grünen v. Facebook Ireland Limited* (Austrian Appellate Court 2017). Austria's Green Party sued Facebook Ireland on behalf of the party's former parliamentary chair and spokesperson, Eva Glawischnig-Piesczek, demanding that the network remove a comment that the party deemed harmful to her reputation. The Austrian Court of Appeal ruled in her favor and demanded that Facebook delete all versions of the comment *worldwide,* thereby rejecting Facebook's argument that it was governed by California law or Irish law, not Austrian law. The case is currently on appeal before the Austrian Supreme Court, which referred questions about the injunction to the CJEU. The CJEU has declared that "EU law does not preclude a host provider such as Facebook from being ordered to remove identical and, in certain circumstances, equivalent comments previously declared to be illegal," and "EU law does not preclude such an injunction from producing effects worldwide, within the framework of the relevant international law which it is for Member States to take into account." CJEU Press Release No. 128/19, Oct. 3, 2019. Available at:
https://curia.europa.eu/jcms/upload/docs/application/pdf/2019-10/cp190128en.pdf

V. VIOLENT EXTREMIST ORGANIZATIONS ONLINE: EXPRESSION AND ASSOCIATION

The growth of the Internet has created new avenues for transnational crime committed by terrorist and extremist organizations. The changing mediums by which crimes are organized and committed pose serious new challenges for international human rights law. With that in mind, this report aims to analyze the activities of terrorist and extremist organizations on social media and the implications of criminalizing these actions for international rights to freedom of expression and association. This report was assembled from legal research across human rights treaties, past decisions by domestic and international courts, domestic legal codes, and other academic articles, primarily restricted to source documents available online in English.

This chapter assesses the different activities and functions of terrorist and extremist groups on social media in the context of transnational crime, including audience development, incitement, information collecting, financing, and online disruption. First, this chapter addresses the tensions between the harm of these individual actions and their propensity for facilitating future crime with their protections under human rights treaties. Second, this chapter analyzes jurisprudence from international and national courts on how criminalizing different categories of extremist activities complies with or violates international rights to freedom of expression and association. Third, this chapter identifies emerging trends in different countries' domestic laws addressing dangerous groups' adaptations to social media. Fourth, this chapter analyzes the policies undertaken by social media companies to combat changing methods of dangerous groups while preserving vigorous community engagement and respecting a diversity of individual speech.

A. Problem Statement: VEO Activities Online

In order to promote their respective causes, violent extremist organizations (VEOs) engage in a large range of activities online, spanning from activities that clearly constitute ordinary crimes—such as hacking or threatening violent attacks—to activities that squarely implicate freedom of expression concerns. These activities generally constitute praise, support or representation for these VEOs or their causes by supporters or sympathizers online.

While propaganda in support of terrorist organizations like ISIS is often treated as an undifferentiated mass, different types of propaganda raise different levels of concern for freedom of expression. Propaganda that aims to promote or otherwise boost support for the organization will often praise and encourage violent and criminal activities, but also praise non-violent, non-criminal activities. For instance, ISIS propaganda can encompass religious guidance that men not shave, or depict ISIS members distributing charity to civilians in the form of food aid.[133] Propaganda that aims to decrease support for the enemy may constitute hate speech or incitement towards the enemy, or alternatively may advance factual and salient condemnations of the enemy government's policies or actions.

This diversity of activities also applies to extremist and dangerous organizations who promote the superiority of one group over another, such as white supremacist groups. There have been greater calls for collective action internationally over the threat posed by such groups in light of attacks in Hanau, Germany and Christchurch, New Zealand by individuals espousing far right extremist philosophies.[134] Different types of praise, support and representation of these causes will also raise different levels of concerns of freedom of expression. While incitement to violence are clear red lines, the line is not so clear if the speech does not reach that level e.g. racist speech.

Given that social media websites have afforded terrorist, extremist and dangerous organizations a host of novel tools to aid their activities, it is helpful to outline exactly what terrorist functions can be carried out via platforms. Platforms' content policies, international tribunal opinions, and regional trends of criminalization indicate the types of activities being carried out on social media websites. The five activities below are a broad amalgamation of these reports and policies, including examples of the tactics used to carry out these activities.

This report focuses on the rights conflicts and case law informing the criminalization of audience development, ideological support, and operational support like collecting information

[133] Graeme Wood, "What ISIS Really Wants," *The Atlantic*, March 2015. Available at: https://www.theatlantic.com/magazine/archive/2015/03/what-isis-really-wants/384980/
[134] Bharath Ganesh, "Jihadis Go to Jail, White Supremacists Go Free," *Foreign Policy*, May 2019. Available at: https://foreignpolicy.com/2019/05/15/jihadis-go-to-jail-neo-nazis-walk-free-christchurch-call-social-media-dignity-digital-hate-culture-tarrant-breivik-bowers-white-supremacists-ardern-macron/

and organizing illegal activities. Though financing and cyberattacks certainly occur online, they appear to come into significantly less conflict with freedoms of expression and association.

1. Audience Development: Identifying, Recruiting, and/or Training Potential Members

Terrorist and extremist organizations target and recruit individuals using social technologies as well as training potential members across the globe via platforms. Often this is done through either generalized publishing of propaganda or individual targeting online, transitioning from public posts on chat rooms or social media platforms to encrypted direct messaging applications.[135]

2. Ideological Support

Social media has also been used to demonstrate ideological support for VEOs, through online activities like sharing propaganda, encouraging future attacks, praising past attacks, praising non-violent acts, and praising ideology/leaders/members. It is an open question whether glorification or praise falls into the legal definition of this category.

Summary statistics from a database maintained by the Global Internet Forum to Counter Terrorism (GIFCT) help characterize the universe of terrorist propaganda online. Of the more than 200,000 unique pieces of content hashed in the database in July 2019, 85.5% were categorized as "Glorification of Terrorist Acts," 9.1% consisted of "Radicalization, Recruitment, Instruction," 4.8% depicted "Graphic Violence Against Defenseless People," and just 0.4% posed an "Imminent Credible Threat."[136]

3. Operational Support: Collecting Information and Coordinating Attacks

Most if not all users of the internet and social media use them as tools for collecting and/or organizing information. Terrorist and extremist organizations utilize innocuous applications and

[135] Brian Fishman, "Crossroads: Counter-Terrorism and the Internet," *Texas National Security Review*, Feb. 2, 2019. Available at: https://tnsr.org/2019/02/crossroads-counter-terrorism-and-the-internet/. Ezekiel Rediker, "The Incitement of Terrorism on the Internet: Legal Standards, Enforcement, and the Role of the European Union," 36 *Mich J. Int'l L.* 321 (2015). Available at: https://repository.law.umich.edu/mjil/vol36/iss2/3.
[136] *GIFCT Transparency Report*, Global Internet Forum to Counter Terrorism (GIFCT), July 2020. Available at: https://www.gifct.org/transparency/

tools such as Google Maps or Google Search to gather information on weapons, attack locations, as well as using social media and direct messaging platforms to coordinate attacks.

4. Financing: Collecting and Transferring Funds

Terrorist and extremist organizations similarly use encrypted messaging, propaganda, and financial transfer tools such as Western Union and cryptocurrency to finance their operations and gather support from donors. Social media can facilitate these transfers of funds.

5. Online Disruption: Cyberattacks

Hacking, doxing, and disrupting networks are some of the many ways that terrorist and extremist organizations can utilize the internet to virtually target users. Doxing consists of posting a target's personal information online. For instance, a terrorist organization might post the home addresses of security personnel, which raises the specter of followers using that information to attack those locations.

B. Expert Guidance on Relevant International Human Rights Standards

While dangerous groups' use of social media can increase the scale and severity of the danger they pose, their online activities are nonetheless subject to protections under international human rights law. In particular, actions on social media and digital participation in online groups implicate freedoms of expression and association.[137]

As described in greater detail in the above chapter on IHRL architecture, many international and regional human rights law mechanisms protect these rights, but also subject them to restrictions to protect national security, public health, and public morality. Reputable organizations have issued interpretive guidance which may help clarify the meaning of these protections and suggest factors that states should consider when implementing restrictions on speech. The pieces of guidance most relevant to violent extremist activity online and therefore highlighted in this section include: the Rabat Plan of Action, which was endorsed by the UN Office

[137] Bart Cammaerts, "Radical pluralism and free speech in online public spaces: the case of North Belgian extreme right discourses," *International Journal of Cultural Studies*, 12 (6), Oct. 2009.

of the High Commissioner for Human Rights; EU Directive 2017/541; and the Johannesburg Principles on National Security, Freedom of Expression and Access to Information, which was adopted by a conference of human rights experts.

1. Rabat Plan of Action

The Rabat Plan of Action, issued by the UN Office of the High Commissioner for Human Rights in 2013, demonstrates modern methods for evaluating hate speech. In particular, the Rabat Plan of Action uses a six-part test for assessing whether an expression constitutes criminal hate speech that rises to the level of inciting violence, including the (1) context of the statement, (2) position/status of the speaker, (3) intent of the statement, (4) content of the statement, (5) extent/spread of the speech act, and (6) the likelihood of incitement.[138] This framework can help better ground the application of existing treaties like the ICCPR, ACHR, and ECHR to the speech of dangerous groups on social media platforms.[139]

2. EU Directive 2017/541

While not binding international law, a 2017 directive from the European Parliament and Council provides helpful definitions for terrorist offences such as public provocation to commit a terrorist act, and training for terrorism. Directives, unlike regulations, are not directly enforceable on a domestic level. However, they do set minimum guidelines and expectations for the practices of each member country. This directive identifies principles that have been cited in European court decisions (see Jurisprudence section 3) and thus provide some substantiated guidance that may be generalizable for European cases, and perhaps even beyond the EU.

First, the directive robustly defines how the court should consider "the offence of public provocation to commit a terrorist offence act . . . the glorification and justification of terrorism or the dissemination of messages or images online and offline . . . as a way to gather support for terrorist causes or to seriously intimidate a population." The Directive urges courts to consider these acts as punishable only when they cause a "danger that terrorist acts may be committed." In

138 "Rabat Plan of Action."
139 *Ibid.*

order to determine whether this danger is present, the courts should take into account the specific circumstances of the case such as: (1) the author and addressee of the message and (2) the context in which the act is committed.[140]

Further, the directive provides specific guidance on what constitutes receiving and providing training, recommending that national law criminalize these offenses based upon the threats that result from this training. This, the directive clarifies, includes those that are "acting alone." Thus, "self-study, including through the internet or consulting other teaching material should also be considered to be receiving training for terrorism when resulting from active conduct and done with the intent to commit or contribute to the commission of a terrorist offence." Again, the directive advises that the context is heavily considered when inferring intent, explaining that "downloading a manual to make explosives for the purpose of committing a terrorist offence" would be receiving training, while "merely visiting websites or collecting materials for academic purposes" is not considered training. The directive recommends generally examining "the type of materials and the frequency of reference" to infer intent.

3. Johannesburg Principles on National Security, Freedom of Expression and Access to Information

The Johannesburg Principles on National Security, Freedom of Expression and Access to Information, standards adopted by a conference of human rights experts in 1995, stipulate further detail on what entails "national security" concerns and "necessary" action when restricting expression.[141]

Under these Principles, "necessity in a democratic society" requires governments restricting speech to conclude that the following three conditions are met:

1. the expression or information at issue poses a serious threat to a legitimate national security interest;

[140] *Directive (EU) 2017/541 of the European Parliament and of the Council of 15 March 2017 on combating terrorism and replacing Council Framework Decision 2002/475/JHA and amending Council Decision 2005/671/JHA*, EUR-Lex, Mar. 15, 2017. Available at: http://data.europa.eu/eli/dir/2017/541/oj

[141] Article 19, *The Johannesburg Principles on National Security, Freedom of Expression and Access to Information*, 1 Oct. 1995, available at: https://www.refworld.org/docid/4653fa1f2.html

2. the restriction imposed is the least restrictive means possible for protecting that interest; and
3. the restriction is compatible with democratic principles.

Further, the Principles suggest expression may be punished as a threat to national security only if a government can demonstrate that:

1. the expression is intended to incite imminent violence;
2. it is likely to incite such violence; and
3. there is a direct and immediate connection between the expression and the likelihood or occurrence of such violence.

These standards have been relied upon by some number of judges, lawyers, and academics.[142] While these Principles are not legally binding, nor have any countries made a formal or informal commitment to the Principles, they serve as another set of criteria similar to the Rabat Plan of Action that should guide decisions to restrict expression.

C. Jurisprudence on International Human Rights

In light of the tension between human rights and the actions of terrorist and extremist groups on social media, examining past cases provides meaningful insight into how these frameworks are currently being interpreted in national and international law. First, the research team analyzed cases on representation: criminal prosecutions of people for declaring themselves supporters of, otherwise displaying membership in, or performing recruitment on behalf of, entities engaging in illegal violence. Second, the team analyzed prosecutions for operational support activities: collecting and transferring funds on behalf of, or coordinating and planning attacks for, such entities. Third, the team analyzed cases involving ideological support: sharing or reposting an entity's propaganda without commentary indicating sentiment, encouraging future attacks, praising past attacks, and praising the entity's goals, ideology, leaders, or members.

[142] Toby Mendel, *The Johannesburg Principles: Overview and Implementation*, Article 19, Feb. 7, 2003. Available at: https://www.article19.org/data/files/pdfs/publications/jo-burg-principles-overview.pdf

This typology encompassed all found international and international cases analyzing the freedom of expression or association implications of prosecutions of activities in support of VEOs. For instance, the research team found no such cases in which people were prosecuted for praising specific non-violent acts of a violent entity, such as ISIS personnel handing out food to poor residents of Mosul.[143] The team's search included all cases from international courts available in English, in addition to a number of cases in which detailed summaries in English could be found. The team also canvassed national court databases available in English, making sure to find at least one on-point national case from each of the following five categories: Europe; Southeast and East Asia; Africa and the Middle East; U.S., U.K., Canada, and Australia; and South America. A number of cases fell into multiple categories of activities; analysis of these cases accordingly re-appears multiple times in the below sections.

In total, the research team found and analyzed twenty-nine cases that examined issues of freedom of expression or association related to VEO activity, whether online or offline. The vast majority related to ideological support: twenty-two cases. Very few cases on operational support mentioned free expression or association: one on funding and one on coordinating attacks. Another four related to membership. Therefore, the conclusions we can draw primarily relate to ideological support cases.

Virtually all court cases cited below articulated the same basic rule for analyzing a prohibition on support for or membership in a VEO, the same rule articulated in the chapter on IHRL Architecture. This basic rule, for either ECHR Article 10 or 11 (or their regional counterparts), was consistently phrased as follows. Accordingly, this rule is articulated here rather than reiterated repeatedly in the case summaries.

The basic three-part test is: in order to justify a restriction on freedom of expression or freedom of association, the restriction must be (1) prescribed by law, (2) pursue a legitimate aim (such as national security, public safety, prevention of disorder, protection of the rights and freedoms of others), and be (3)(a) necessary in a democratic society as well as (3)(b) proportionate.

[143] Hamza Hendawi and Bassem Mroue, "ISIS is offering a mix of brutality and charity during Ramadan," *The Business Insider*, July 10, 2015. Available at: https://www.businessinsider.com/isis-is-offering-a-mix-of-brutality-and-charity-during-ramadan-2015-7

The substance of courts' analysis almost always lies in analyzing the necessity and proportionality of the restriction, so it is here where the detailed case analyses—as well as the below framework of analysis—focuses.

The exception is for the ECtHR, which twice, in cases we analyzed, applied ECHR Article 17, "Prohibition of abuse of rights," instead of ECHR Article 10 or 11 frameworks for restrictions on freedom of expression or association.[144] ECHR Article 17 declares that actions "aimed at the destruction of any of the rights and freedoms set forth" elsewhere in the ECHR are ineligible for the protections of the ECHR. According to the ECtHR, the purpose of Article 17 is to prevent totalitarian organizations from exploiting the freedoms enshrined in the ECHR. For instance, when a speaker directs a remark "against the Convention's underlying values," he forfeits eligibility for Article 10 protection of free expression.[145] The only two cases identified in which the ECtHR applied Article 17 instead of Article 10 or 11 involved a state's decision to shutter an entire organization, rather than prosecutions of individuals. Specifically, the ECtHR used Article 17 to uphold the banning of an Islamist organization calling for the violent destruction of Israel[146] and the fining and revocation of the broadcasting license of a TV station broadcasting PKK messages.[147]

1. Analytical Framework

Based on these twenty-nine cases, we have developed a framework of eight key factors that courts consider when determining whether or not a restriction on free speech is justified, i.e. whether or not it is necessary in a democratic society and proportionate. None of the courts we analyzed considered all of these factors in any one case, but all cases we found that analyzed freedom of expression or association mentioned at least two of these factors.

144 Full Text of Article 17: "Nothing in this Convention may be interpreted as implying for any State, group or person any right to engage in any activity or perform any act aimed at the destruction of any of the rights and freedoms set forth herein or at their limitation to a greater extent than is provided for in the Convention." ECHR.
145 *Hizb ut-Tahrir and Others v. Germany*, ¶ 72 (Citing *Paksas v. Lithuania* [GC], no. 34932/04, § 45, §§ 87-88, 6 2011).
146 *Hizb ut-Tahrir and Others v. Germany*, No. 31098/08 (ECtHR 2012).
147 *Roj TV A/S v. Denmark*, No. 24683/14, (ECtHR, Apr. 17, 2008).

This framework resembles guidelines like the Rabat Plan of Action and Johannesburg Principles, but is not derived from them. Rather, all these typologies of factors to be considered follow the inherent logic of how to weigh both content and context of a prosecuted post.

(1) Nature of the post
(2) Number of the post(s)
(3) Content of the post
(4) Resulting violence, if any
(5) Timing of the post
(6) Medium and reach
(7) Speaker's role and personal history
(8) Proportionality of the sentence

The first three of these factors relate to the post itself: (1) the nature and (2) number of the post(s) and (3) content of the post. Glorification or praise of violence may constitute incitement, and therefore justify restrictions on free expression (including criminalization), when the content endorses violence. The next is (4) whether or not any actual violence resulted from the post. The next three elements of analysis relate to the real-world context of the post: (5) the timing of the post, (6) the medium and reach of the post, and (7) the role and personal history of the speaker. The final factor is (8) the proportionality of the sentence imposed on the speaker.

In practice, courts varied in how they weighed these elements against each other. Many courts, in order to uphold criminalization of speech, required the content of the post to explicitly advocate violence—even when contextual factors weighed heavily in favor of restriction.[148]

However, other courts were willing to uphold criminalization even if a post had only implicitly endorsed violence (such as by ambiguous but mildly positive references to former attacks). These courts weighed contextual factors more heavily than content-based ones.[149]

Various categories of content never, in the cases we analyzed, justified criminal prosecution resulting in sentences of imprisonment: (1) criticizing the government's suppression

[148] E.g., *Hussain v. The Norwegian Prosecution Authority*, Nos. 14-0499903AST-BORG/01, 14-174730AST-BORG/01 (Norwegian Appellate Court, June 22, 2015)
[149] E.g., *Zana v. Turkey*, No. 40984/07 (ECHR Grand Chamber, Nov. 25, 1997).

of a VEO;[150] (2) praising a VEO's goals or ideology without endorsing its violent methods;[151] and (3) praising a VEO's leader without referencing his violent or criminal acts.[152] This finding is discussed in more detail under (3) Content of the Post.

(1) Nature of the Post(s)

Courts consider whether a post is intended or appears to be satire rather than sincere advocacy,[153] or is phrased hypothetically.[154,155] Even a clear satirical or joking stance, however, does not always save a post from being considered to justify restriction on free expression.[156]

(2) Number of the Post(s)

Courts also consider the number of complained-of posts. If a speaker or entity has engaged in a large number of provocative statements,[157,158,159] courts are more likely to ascribe intent to endorse or encourage violence than if the statements complained of are few in number.[160]

(3) Content of the Post

The single most important factor—and sometimes the only one to which courts pay attention, other than proportionality of the sentence—is what the post says. Courts consider whether the post is specific or ambiguous in its support for a VEO or its violent activities.

When only one remark within a lengthy speech or post appears to support violent extremist activity, courts often consider the text surrounding the remark and text found elsewhere in the piece. For instance, courts ask whether the sentences alleged to incite violence are aberrant in the

150 E.g., *The Case of Ayse Celikel*, App. No. 2017/36722 (Turkish Constitutional Court, May 9, 2019).
151 E.g., *The Case of the Academics for Peace*, No. 2018/17635 (Turkish Constitutional Court, Jul. 26, 2019).
152 E.g., *Faruk Temel v. Turkey*, No. 45281/08 (ECtHR Second Section, Apr. 24, 2018).
153 *The State v. Cassandra Vera*, STC 493/2018 (Spanish Supreme Court, Feb. 26, 2018) .
154 *Smajić v. Bosnia and Herzegovina*, No. 48657/2016 (ECtHR 2018).
155 *Fatullayev v. Azerbaijan*, No. 40984/07, (ECtHR 2010).
156 *Leroy v. France*, Legal Summary, No. 36109/03 (ECHR, Feb. 2, 2008).
157 *Kaptan v. Switzerland*, No. 55641/00 (ECtHR Second Section, Apr. 12, 2001).
158 *Hizb ut-Tahrir and Others v. Germany*, No. 31098/08 (ECtHR 2012).
159 *R. v. Iqbal*, 2014/01692 C5 (EWCA Crim. 2650, 2014)
160 *Mahajna v. Home Secretary*, UKUT B1 (IAC) (Upper Tribunal, Immigration and Asylum Chamber, Apr. 16, 2012).

larger context of the piece or mediated by more pacific sentences elsewhere in the post, versus if they appear to represent the post's central message or thesis.[161]

There is a spectrum of likelihood that courts will find a restriction, especially criminalization, justified based on the content of a piece of speech. Justification is very likely for content that explicitly encourages violence,[162] compared to moderately likely for content that explicitly praises past attacks.[163] Justification is considerably less likely for content that only implicitly justifies or approves of violence,[164] such as statements which joke about or ambiguously refer to attacks, without defending them as justified.[165] In these borderline cases, some courts weigh contextual factors more heavily. Others will refuse to weigh contextual factors when the content does not clearly communicate approval of violence. Even omissions (such as failures to condemn attacks by others) have sometimes been considered to justify restriction,[166,167] although this represents the exception rather than the general rule.[168]

As a result, the following functions for a post did not justify restriction in any of the cases we identified: sharing a VEO's non-violent political goals or supporting its underlying ideology[169,170,171] or criticizing government actions or policies.[172,173] Courts emphasize that criticism of the government on matters of public interest is entitled to particularly strong protection as free expression, and that the restraint governments must show in criminalizing speech makes imprisonment a wholly disproportionate response to criticism of public authorities.[174,175,176]

161 *Ibid.*
162 *Hizb ut-Tahrir and Others v. Germany*, No. 31098/08 (ECtHR 2012).
163 *Case of Jose Miguel Arenas*, No. 79/2018 (Spanish Supreme Court, Feb. 15, 2018).
164 *The State v. Cassandra Vera*, STC 493/2018 (Spanish Supreme Court, Feb. 26, 2018).
165 E.g., *The State v. Cassandra Vera*, STC 493/2018, (Spanish Supreme Court, Feb. 26, 2018).
166 *Herri Batasuna and Batasuna v. Spain*, Nos. 25803/04 and 25817/04 (ECtHR 2009).
167 *Leroy v. France*, Legal Summary, No. 36109/03 (ECHR, Feb. 2, 2008).
168 *Lehideux and Isorni v. France*, No. 24662/94 (ECHR Grand Chamber, Sept. 23, 1998).
169 *The Case of the Academics for Peace*, No. 2018/17635 (Turkish Constitutional Court, Jul. 26, 2019).
170 *The Case of Ayse Celikel*, App. No. 2017/36722 (Turkish Constitutional Court, May 9, 2019).
171 *Keun-Tae Kim v. Republic of Korea*, No. 574/1994 (UN Human Rights Committee, Nov. 3, 1998).
172 *Fatih Tas v. Turkey* (No. 3), No. 45281/08 (ECHR Grand Chamber, Apr. 24, 2018).
173 *Faruk Temel v. Turkey*, No. 45281/08 (ECtHR Second Section, Apr. 24, 2018).
174 *Ibid.*
175 *The Case of the Academics for Peace*, No. 2018/17635 (Turkish Constitutional Court, Jul. 26, 2019).
176 *The Case of Ayse Celikel*, App. No. 2017/36722 (Turkish Constitutional Court, May 9, 2019) [in Turkish].

(4) Any Violence Resulting from the Post

A post is more likely to justify restriction if violence appeared to result from the post,[177] and less likely if no violence resulted.[178,179] However, no court required that actual violence result from a post in order to deem a post incitement and justify restriction.[180]

As correlation does not equal causation, courts did not necessarily assume that violence following a post proved that the post constituted incitement and therefore justified restriction. For instance, when violent protests broke out after a party official distributed pamphlets, the UN HRC held his conviction for distributing the pamphlets nevertheless violated free expression because the pamphlets advocated only peaceful actions.[181] In another instance, a court dismissed an allegation that a speech had led to a nearby riot, noting that a video of the speech itself showed no violence.[182]

(5) Timing of the Post

A post is more likely to justify restriction if it is posted very soon after a VEO attack,[183] or during a very sensitive political situation, such as ongoing violent unrest.[184, 185] It is less likely to justify restriction if it apparently endorses an attack or crime perpetrated decades ago.[186,187]

(6) Medium & Reach of the Post

Courts sometimes consider the medium of the content: whether support is being expressed via a newspaper, blogging website, social media.[188] The more public the platform, the more likely the speech is to justify restriction, based on the ability of the posts to reach mass audiences.[189]

177 *Keun-Tae Kim v. Republic of Korea*, No. 574/1994 (UN Human Rights Committee, Nov. 3, 1998)
178 *Mahajna v. Home Secretary*, UKUT B1 (IAC) (U.K. Upper Tribunal, Immigration and Asylum Chamber 2012).
179 *Vanjai v. Hungary*, No. 33629/06 (ECtHR, July 8, 2008).
180 *Case of Jose Miguel Arenas*, No. 79/2018 (Spanish Supreme Court, Feb. 15, 2018).
181 *Keun-Tae Kim v. Republic of Korea*, No. 574/1994 (UN Human Rights Committee, Nov. 3, 1998).
182 *Mahajna v. Home Secretary*, UKUT B1 (IAC) (U.K. Upper Tribunal, Immigration and Asylum Chamber 2012).
183 *Leroy v. France*, Legal Summary, No. 36109/03 (ECHR, Feb. 2, 2008)
184 *The State v. Cassandra Vera*, STC 493/2018 (Spanish Supreme Court, Feb. 26, 2018).
185 *Zana v. Turkey*, No. 40984/07 (ECHR Grand Chamber, Nov. 25, 1997).
186 *Lehideux and Isorni v. France*, No. 24662/94 (ECHR Grand Chamber, Sept. 23, 1998).
187 *The State v. Cassandra Vera*, STC 493/2018 (Spanish Supreme Court, Feb. 26, 2018).
188 *Smajić v. Bosnia and Herzegovina*, No. 48657/2016, (ECtHR 2018).
189 *The State v. Cassandra Vera*, STC 493/2018 (Spanish Supreme Court, Feb. 26, 2018).

Editorials or articles, deliberately written and selected for publication,[190] might be more likely to support restriction than hasty spur-of-the-moment speech, such as a call-in comment to a broadcast.[191]

(7) Speaker's Role and Personal History

Courts often consider a speaker's role, particularly in terms of his or her ability to exert authority or influence, e.g. as a leader of a political party,[192] former office-holder,[193] leader of an extremist group, or editor,[194] versus that of an academic,[195] private citizen,[196] or satirist.[197] Speech from a speaker with great prominence or a large number of followers is more likely to justify restriction.[198, 199]

Courts may also consider a speaker's prior behavior or personal history and circumstances when determining whether to ascribe knowledge or intent to incite violence to a piece of speech.[200, 201, 202]

(8) Proportionality of Sentencing

Sentences of two years or longer in prison for ideological support offenses were almost uniformly held to violate freedom of expression,[203] especially when the speakers did not explicitly incite violence.[204] Only two courts upheld such prison sentences; a U.K. case involving

190 *Fatih Tas v. Turkey* (No. 3), No. 45281/08 (ECHR Grand Chamber, Apr. 24, 2018).
191 *The Case of Ayse Celikel*, App. No. 2017/36722 (Turkish Constitutional Court, May 9, 2019) [in Turkish].
192 *Keun-Tae Kim v. Republic of Korea*, No. 574/1994 (UN Human Rights Committee, Nov. 3, 1998).
193 *Zana v. Turkey*, No. 40984/07 (ECHR Grand Chamber, Nov. 25, 1997).
194 *Fatih Tas v. Turkey* (No. 3), No. 45281/08 (ECHR Grand Chamber, Apr. 24, 2018).
195 *The Case of the Academics for Peace*, No. 2018/17635 (Turkish Constitutional Court, Jul. 26, 2019).
196 *Fatullayev v. Azerbaijan,* No. 40984/07 (ECtHR 2010).
197 *Tribunal de grande instance de Paris (jugement correctionnel du 18 mars 2015)* (High Court of Paris, Mar. 18, 2015).
198 *Ibid.*
199 *Zana v. Turkey*, No. 40984/07 (ECHR Grand Chamber, Nov. 25, 1997).
200 *The Case of Dmitry Semenov*, 22-2559/2015, (Russian Appellate Ct. 2015).
201 *Mahajna v. Home Secretary, UKUT B1 (IAC) (U.K. Upper Tribunal, Immigration and Asylum Chamber 2012).*
202 *The State v. Cassandra Vera*, STC 493/2018 (Spanish Supreme Court, Feb. 26, 2018).
203 E.g., *Stomakhin v. Russia*, No. 52273/07 (ECtHR 2017).
204 E.g., *Fatullayev v. Azerbaijan,* No. 40984/07 (ECtHR 2010).

encouragement that people travel to join ISIS,[205] and a Spanish case involved explicit justifications of violence and explicit praise of past violence.[206]

Even shorter prison or jail sentences for speech supporting VEOs were usually considered to violate freedom of expression.[207] Out of twenty-eight total cases, eleven involved sentences of imprisonment. Only three upheld these sentences as consistent with free expression,[208,209,210] The remaining eight held the prison sentences violated free expression.[211]

Two cases involved suspended prison sentences. One found the sentence proportionate for ideological support that fell short of incitement to violence,[212] while the other did not.[213]

The remaining cases involved civil penalties or criminal fines, which often did not violate free expression. Eight held the restrictions justified,[214] compared to six that found violations.[215]

Courts apply the eight key factors identified in this analytic framework when assessing the necessity and proportionality of a restriction on free expression or association. The following sections summarize and analyze a survey of twenty-eight international and national court cases analyzing free speech and association concerns associated with restricting support of VEOs. This section also highlight a number of specified "national analogues:" cases we came across in which courts analyzed convictions for terrorist speech under national legal principles, without mentioning free expression or association.

205 *Choudary & Anor v. Regina* (England and Wales Court of Appeal, Criminal Division, Mar 22, 2016).
206 *Case of Jose Miguel Arenas*, No. 79/2018, (Spanish Supreme Court, Feb. 15, 2018).
207 *The State v. Cassandra Vera,* STC 493/2018, (Spanish Supreme Court, Feb. 26, 2018).
208 *Choudary & Anor v. Regina*, England and Wales Court of Appeal (Criminal Division) (Mar 22, 2016).
209 *Case of Jose Miguel Arenas*, No. 79/2018, (Spanish Supreme Court, Feb. 15, 2018).
210 *Zana v. Turkey*, No. 40984/07 (ECHR Grand Chamber, Nov. 25, 1997).
211 E.g. *Fatullayev v. Azerbaijan*, No. 40984/07 (ECtHR 2010).
212 *Tribunal de grande instance de Paris (jugement correctionnel du 18 mars 2015)* (High Court of Paris, Mar. 18, 2015).
213 *Smajić v. Bosnia and Herzegovina*, No. 48657/2016 (ECtHR 2018).
214 E.g. *Case of Saygili and Falakaoglu v. Turkey* (No. 2), No. 38991/02 (ECtHR Grand Chamber, Feb. 27, 2009); *R. v. Iqbal*, 2014/01692 C5 (EWCA Crim. 2650, 2014); *Leroy v. France, Legal Summary*, No. 36109/03 (ECHR, Feb. 2, 2008).
215 E.g. Granier et al. v. Venezeula, Report on Merits, Report No. 112/12, Case No. 12.828 (Inter-Am. Comm'n H.R., Nov. 9, 2012); *Vanjai v. Hungary*, No. 33629/06 (ECtHR, July 8, 2008).

2. Representation

Representation is defined here as belonging to an organization or calling for others to join it. The first category of cases concerns the permissibility of banning organizations themselves, analyzed at the aggregate level of the entire association's right to exist rather than prosecutions for individual members. The second category deals with prosecutions of individuals for recruiting others to join a banned organization.

a. Membership and Association

Two primary cases surfaced that assessed freedom of association with reference to banning particular organizations as a result of, in part, their online activities. In both, the ECtHR held the dissolution or banning of the organizations was consistent with freedom of association. The first held an Islamic organization ineligible for the protection of freedom of association under ECHR Article 17 because one of the group's main goals was killing the people of Israel, which is incompatible with the fundamental values of the ECHR.[216] The second held the interference with two political parties' freedom of expression justified due to their organizational links to and failure to disavow the violence of a Basque separatist terrorist organization. [217]

A third case, from an Italian national court, assessed a constitutional analogue to international freedoms of association and expression, with respect to an organization's ability to speak online. It found Facebook's banning of a political party unjustified; Facebook had failed to show a link between statements properly ascribed to the party and resultant violence.[218]

(i) Organization's Website, Offline Materials: Organization Banned – Free Association and Expression Not Violated

Case: Hizb ut-Tahrir and Others v. Germany, ECtHR, 2012.

Synopsis: In 2012, the ECtHR held that under Article 17, Islamic organization Hizb Ut-Tahrir's may be banned; its activities were unprotected by Articles 9, 10, or 11 because its aims

[216] *Hizb ut-Tahrir and Others v. Germany*, No. 31098/08 (ECtHR 2012).
[217] *Herri Batasuna and Batasuna v. Spain*, Nos. 25803/04 and 25817/04 (ECtHR 2009).
[218] Full text in Italian is available at https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2020/01/sentenzacpifb.pdf

were incompatible with the values of the ECHR.[219] The ECtHR deemed that, as demonstrated by many statements attributable to Hizb Ut-Tahrir, one of the association's "main concerns" was "call[ing] for the violent destruction of [Israel] and for the banishment and killing of its inhabitants."[220]

In 2003, German authorities banned and confiscated the assets of an organization called Hizb Ut-Tahrir (Party of Liberation), on the basis that this association advocated violence in order to achieve political goals.[221] Hizb Ut-Tahrir's activities in Germany included distributing online propaganda and offline leaflets and brochures, and organizing in-person events.[222] In particular, the German ministry argued that articles published in the organization's quarterly magazine *Explizit* and publications on the organization's website "denied the right of the State of Israel to exist and called for its destruction and for the killing of Jews" as well as repeatedly calling for the overthrow of various Islamic governments.[223]

For instance, one 2002 *Explizit* article argued that Israel could not be allowed to continue to exist, that jihad was the only possible response to "Zionist aggression, and that "Allah, the Exalted, commands: 'And slay them wherever ye catch them, and turn them out from where they have turned you out.'"[224] A 2001 Hizb Ut-Tahrir flyer read: "Muslims are duty bound to liberate [Palestine] from the rule of the Jews, even if it costs the lives of millions of martyrs."[225] Another repeated the "slay them wherever ye catch them" verse from the Qur'an while advocating: "The solution is to uproot the Jewish Entity from the entire Palestine . . . every negotiation with the Jews is treason against Allah."[226] Hizb Ut-Tahrir's representative in court, Hussein Assem Shaker, justified suicide attacks in Israel as "self-defence," with his words broadcast on Berlin local television in 2002. He argued that while Islam forbids violence against civilians, "there are no

[219] *Hizb ut-Tahrir and Others v. Germany*, No. 31098/08 (ECtHR 2012).
[220] *Ibid*.
[221] The German Law on Associations permits banning an association if "its aims or its activities contravene the criminal law or . . . are directed against the constitutional order or against the idea of international understanding." *Ibid*.
[222] *Ibid*.
[223] *Ibid*, ¶ 6.
[224] *Ibid*, ¶ 15.
[225] *Ibid*, ¶ 22.
[226] *Ibid*, ¶ 22.

civilians in Israel" because all Israelis "are part of the military," that everyone who lives there is complicit in "the act of aggression" which founded Israel, and that "[i]f children are also killed, their parents are responsible for having decided to live in Israel."[227]

A German court found the prohibition proportionate, reasoning that Hizb Ut-Tahrir was not entitled to special license as a religious community because of its primarily political aims, but even if they were, they could still be prohibited because of its "multitude of public statements" calling for the violent deaths of Israelis.[228] Hizb Ut-Tahrir, as well as a group of its representatives, members, and supporters, applied to the ECtHR, alleging violations of their rights to free expression, association, and religion under ECHR Articles 9, 10, and 11.[229]

The ECHR held that under Article 17, Hizb Ut-Tahrir's activities were unprotected by Articles 9, 10, or 11 because, as demonstrated by many statements attributable to Hizb Ut-Tahrir, one of the organization's "main concerns" was "call[ing] for the violent destruction of [Israel] and for the banishment and killing of its inhabitants."[230] The ECHR first recapitulated its Article 17 jurisprudence, emphasizing that Article 17's purpose was "to prevent individuals or groups with totalitarian aims from exploiting in their own interests the principles enunciated in the Convention." Article 17 operates by removing from the protection of Article 10 or Article 11 for any "remark directed against the Convention's underlying values."[231] The ECHR had previously held Article 17 barred an anti-Semitic association from relying on Article 11 rights to free association to challenge its prohibition.[232]

Applying this doctrine to Hizb Ut-Tahrir, the ECHR agreed with German courts that Hizb Ut-Tahrir's aims were contrary to the values of the Convention. This conclusion was based on articles published by Hizb Ut-Tahrir and public statements by its representative in court, which "repeatedly justified suicide attacks in which civilians were killed in Israel."[233] The court

[227] *Ibid*, ¶ 21.
[228] *Ibid*, ¶ 25–28.
[229] *Ibid*, ¶ 35.
[230] *Ibid*, ¶ 78.
[231] *Ibid*, ¶ 72 (Citing *Paksas v. Lithuania* [GC], no. 34932/04, § 45, §§ 87-88, 6 Jan. 2011).
[232] *Ibid*, ¶ 72.
[233] *Ibid*, ¶ 73.

emphasized that Hizb Ut-Tahrir's articles and flyers not only "called for the violent destruction of [Israel] and for the banishment and killing of its inhabitants," but that these aims were one of its "main concerns" as an organization, and neither the organization or its court representative had distanced themselves from this goal during the ECHR proceedings.[234] Accordingly, Article 17 barred the organization's reliance on the protection of freedom of association under Article 11, freedom of expression under Article 10, or freedom of religion under Article 9.

(ii) Organization's Website, Statements: Parties Banned – Free Association Not Violated

Case: *Herri Batasuna and Batasuna v. Spain*, ECtHR, 2009.

Synopsis: In 2009, the ECtHR held the dissolution of the two applicant Basque parties by the Spanish government was a necessary and proportionate restriction on their Article 11 rights to free association, due to the parties' organizational links to and failure to disavow the violence of a terrorist organization, Euskadi Ta Askatasuna (ETA).[235] The court considered the conduct of the party and its leaders, including many statements tacitly supportive of ETA's violence, as well as their omissions (failing to condemn ETA's violence) in determining their party activities incompatible with a democratic society. The court also considered the political sensitivity of the Basque region in finding a plausible threat to public order posed by the dissolved parties and the confrontational climate to which they contributed.

In 2002, Spanish authorities sought court orders dissolving three left-wing Basque separatist political parties: Herri Batasuna, Euskal Herria (EH) and Batasuna.[236] In 2003, the Spanish Supreme Court declared the parties illegal on grounds that they were jointly controlled by and had defended rather than disavowing violence committed by the terrorist organization Euskadi Ta Askatasuna (ETA).[237] For instance, in 2002, a spokesperson for Batasuna "didn't want ETA to stop killing, but did not want Euskal Herria [EH] to have recourse to any kind of violence and wanted those who engaged in it to cease to exist."[238] Under Spanish law, a political party may only

[234] *Ibid*, ¶ 73.
[235] *Herri Batasuna and Batasuna v. Spain*, Nos. 25803/04 and 25817/04 (ECtHR 2009).
[236] *Ibid*, ¶ 30.
[237] *Ibid*, ¶ 27–28.
[238] *Ibid*, ¶ 34.

be dissolved "in the event of repeated or accumulated acts which unequivocally prove the existence of undemocratic conduct."[239] In finding the parties controlled by the ETA, Spanish courts considered that an ETA delegate had controlled the selection and appointment process for leadership of the three parties, which all shared personnel, including a joint spokesperson.[240] Two of the dissolved parties applied to the ECtHR for relief.

In finding the interference with the parties' freedom of association "prescribed by law," the ECtHR dismissed the parties' allegation that the law was being applied retroactively; the parties took the key complained-of actions after the law went into effect. However, the court noted the ECHR lacks a provision forbidding reliance on facts preceding a law's enactment in order to analyze the justifiability of a restriction on freedom of association.[241]

The ECtHR found several legitimate aims motivated the dissolution, including the protection of public safety, disorder, and the rights of others.[242] The court dismissed the parties' argument that the Spanish government was attempting to suppress the views of left-wing Basque independence, noting that several separatist parties continue to operate peacefully in Spain.[243]

The court articulated its Article 11 jurisprudence related to necessity in a democratic society. The court noted that dissolution of a political party is a drastic measure and emphasized the need to strictly construe Article 11's exceptions, clarifying that states have "only a limited margin of appreciation" when determining if a restriction is necessary in a democratic society and noting that.[244] In order to benefit from Article 11's protections, a political party's goals must be compatible with fundamental democratic principles and its means of achieving those goals must be legal, democratic, and peaceful.[245] To determine a party's objectives and intentions as a whole, a court must compare its constitution and platform to the actions taken and positions defended by

[239] *Ibid*, ¶ 12 (Citing *Institutional Law no. 6/2002 on political parties*, Ley Orgánica 6/2002 de Partidos Políticos – "the LOPP", June 27, 2002). Under this distinction between aims and conduct, while undemocratic aims may be advanced via democratic means, parties may not base their political activity in "violence, political support for terrorist organisations or violation of the rights of citizens or democratic principles."
[240] *Ibid*, ¶ 32–34.
[241] *Ibid*, ¶ 59.
[242] *Ibid*, ¶ 64.
[243] *Ibid*, ¶ 63.
[244] *Ibid*, ¶ 77–78.
[245] *Ibid*, ¶ 79.

the party's members and leaders.[246] However, if the danger to democracy is well-established and imminent, a state need not wait for a political party to implement an undemocratic policy, but can act preemptively under its positive obligations to secure freedoms under Article 1 of the ECHR.[247] For the restriction to qualify as necessary, there must be plausible evidence that (1) the risk to democracy is "sufficiently and reasonably imminent" and that (2) the party's vision for society was clearly incompatible with democracy, as demonstrated "by the acts and speeches imputable to the political party" taken as a whole.[248]

The court, assuming the two applicant parties constituted a single entity, found the criteria met because the parties implicitly supported ETA's terrorist activities. The court pointed to a Batasuna-organized demonstration with slogans supporting ETA prisoners and threatening phrases such as "long live ETA military." In 2002, a Batasuna representative defended ETA's use of force, saying "ETA [did] not support armed struggle for the fun of it, but that [it was] an organisation conscious of the need to use every means possible to confront the State." Lastly, Batasuna had posted an anagram of an illegal terrorist organization ("Gestoras Pro-Amnistía") on its website."[249] The court assessed this conduct as bearing a "strong resemblance to explicit support for violence" and "commendation of" terrorists and creating a "climate of confrontation" which risked public disorder.[250] The court found that, in light of the political sensitivity of the Basque Country, the link between the parties and ETA could constitute a threat to democracy.

The ECtHR acknowledged the dissolution was based in part on the parties' failure to condemn others' violent actions, but agreed with the Spanish court that their "refusal to condemn violence against a backdrop of terrorism that had been in place for more than thirty years and condemned by all the other political parties amounted to tacit support for terrorism."[251] The court also pointed out the dissolution did not result solely from the parties' omissions, but also from its actions and speeches suggesting active accommodation with terrorist actors. Therefore, ECtHR

[246] *Ibid*, ¶ 80.
[247] *Ibid*, ¶ 81.
[248] *Ibid*, ¶ 83.
[249] *Ibid*, ¶ 85.
[250] *Ibid*, ¶ 86.
[251] *Ibid*, ¶ 88.

found the parties had assisted and politically supported terrorist organizations with aims to overthrow the constitutional order, and their dissolution was necessary in a democratic society as well as proportionate.[252]

(iii) (National Analogue) Facebook Page: Dangerous Organization – Violated Free Political Thought

Case: CasaPound v. Facebook, Court of Rome, 2019.

Synopsis: In this case, an Italian court relied on a national constitutional right of political parties to free political debate, without referencing international human rights, to order that Facebook restore an Italian political party's Facebook page. The court reasoned that Facebook had not adequately proved a link between statements properly ascribed to the party and resultant violence.[253]

In 2019, Facebook deactivated the Facebook accounts of Italian neo-fascist party CasaPound and its administrator, on grounds of hate speech and incitement to violence.[254] An Italian court ruled that this deactivation violated the right to free political debate in the Italian Constitution.[255] In 2019, Facebook Ireland deactivated the Facebook page of CasaPound Italia, a far-right political party, as well as those of its representatives and supporters. After CasaPound requested reactivation, Facebook explained that CasaPound had violated its Terms of Service and Community Standards by posting hate speech and incitements to violence. Facebook had determined CasaPound was a "dangerous organization" and "hate group" on the basis of online content as well as offline factors including protests with Nazi salutes and physical assaults on minorities. CasaPound sought a preliminary injunction from the Court of Rome ordering Facebook to reactivate its page and that of its administrator.

252 *Ibid*, ¶ 91–93.
253 Full text in Italian is available at https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2020/01/sentenzacpifb.pdf
254 *CasaPound v. Facebook*, R. G. 59264/2019 (Court of Rome, Dec. 12, 2019).
255 Article 49, Italian Constitution: "All citizens have the right to freely form political parties in order to contribute by democratic means to national policy."

The Italian court granted the injunction, finding Facebook was bound to uphold national constitutional principles of political pluralism until a violation of its Terms of Services was proved; otherwise, this political association would be unable to effectively express its political positions to the public. The court found that Facebook had erred by focusing on Casapound's activities outside of Facebook without proving a causal link between violence perpetrated by CasaPound members and the indicated online content. The court also objected to ascribing all violent acts of or words by CasaPound members to CasaPound as an organization. The court emphasized that CasaPound had been a legitimate actor in Italian politics for a decade.

Note: No English translation of this decision is available, precluding more detailed analysis. The Italian original also appears to lack details in terms of what content was alleged to be incitement. [256]

b. Recruitment

One case identified assessed freedom of expression as it pertained to recruitment for terrorist or extremist organizations. A British appellate court held the conviction and lengthy imprisonment of two nationals for inviting support for ISIS did not violate freedom of expression under the ECHR.[257] While acknowledging the defendants did not advocate that others commit any violence, the court deemed criminalizing inviting even "intellectual support" for a proscribed organization justified and proportionate, as a necessary incident to effectively banning the organization itself.[258]

Other national court cases, discussed below, criminalized VEO recruitment without reference to free expression or association.

(i) Online Oaths of Allegiance: Inviting Support – Free Expression Not Violated

Case: Choudary & Anor v. Regina, England and Wales Court of Appeal, 2016.

[256] Full text in Italian is available at https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2020/01/sentenzacpifb.pdf
[257] *Choudary & Anor v. Regina* (England and Wales Court of Appeal, Criminal Division, Mar 22, 2016).
[258] *Ibid*, ¶ 46, 50.

Synopsis: In 2016, a British appellate court upheld the conviction of two British nationals for inviting support for ISIS, even though they had advocated for travelling to the Islamic State rather than encouraging violence itself. The court acknowledged the defendants had not incited violence, but deliberately decided not to follow the oft-cited rule of international human rights law that speech cannot be criminalized unless it constitutes incitement, arguing that ECtHR itself had blurred this rule on multiple occasions. However, the court distinguished between advocacy for supporting a proscribed organization and advocacy of an idea which the organization happens to share when interpreting the statute to prohibit intellectual support as well as operational support.

In 2016, a British trial court convicted British nationals Anjem Choudary and Mohamed Rahman of inviting support for a proscribed organization: ISIS.[259] Anjem Choudary co-founded an Islamist organization called al-Muhajiroun subsequently banned by the UK. In 2014, on social media, the two posted oaths of allegiance to ISIS and its leader and posted talks encouraging travelling to the Islamic State (making "hijra"), without specifically urging violence. The posted oath included the exhortation: "Every one of the Muhajiroun has a duty . . . to take jihad to call for Islam and to support the pledge of allegiance to the Caliphate State."[260] The trial judge sentenced them each to five and a half years imprisonment.[261]

On appeal, the defendants argued the trial judge's interpretation of "inviting support" (as not restricted to a practical or tangible form or support) was incompatible with ECHR Articles 7, 9, and 10, as well as Articles 11 and 12 of the Charter of Fundamental Rights and Freedoms of the European Union.[262] The appellate court articulated the usual elements of a valid restriction on free expression under ECHR Article 10: prescribed by law, legitimate aim, necessary in a democratic society, and proportionate.[263] The court found the analysis for rights implicated in the EU Charter identical to that for ECHR Article 10.[264]

[259] U.K. Terrorism Act 2000, § 12(1)(a).
[260] *Choudary & Anor v. Regina*, ¶ 14.
[261] *R v Anjem Choudary and Mohammed Rahman*, England Central Criminal Court (Sentencing remarks of J. Holyroyde) (Sept 6, 2016). Available at: https://www.judiciary.uk/wp-content/uploads/2016/09/r-v-choudary-sentencing.pdf
[262] *Choudary & Anor v. Regina*, ¶ 6–7.
[263] *Ibid*, ¶ 62–67.
[264] *Ibid*, ¶ 90.

In finding the "inviting support" provision properly prescribed by law, appellate court agreed with the trial judge that the type of invited "support" prohibited by the law includes not only operational support, but also "intellectual support," such as advocacy, approval, or endorsement.[265] The court insisted the "inviting support" provision was not too vague for fair notice and legality under ECHR Article 7. However, the court agreed with the trial judge that inviting support must be distinguished from merely expressing a personal belief which the proscribed organization happens to share, "or an invitation to someone else to share [that] opinion or belief."[266]

The court briefly noted that the statute at hand was "clearly directed" to the legitimate aims of national security, public safety, public order, and the rights and freedoms of others, before analyzing necessity.[267] The court declared that banning invitations of support for proscribed organizations—even intellectual rather than operational support—is essential for effectively banning the organizations themselves. The court reasoned that a group with more supporters will grow stronger and more determined.[268] When finding the interference proportionate and justified, the court emphasized that the statute required knowledge, prohibiting only the "knowing invitation of support from others" for a proscribed organization.[269]

The appellate court characterized "Strasbourg" jurisprudence on Article 10 as posing a "bright line" rule that states may only criminalize speech if it "advocates or encourages violence." Furthermore, the court acknowledged that the defendants' posted statements contained no explicit invitations to violence. However, the court went on to profess itself unpersuaded that any "bright line" principle could be discerned from the variety of ECtHR cases on Turkish prosecutions for sharing propaganda, and pointed out that none of those involved inviting support from third parties for the proscribed organization.[270] The court also noted two previous UK cases had held

[265] *Ibid*, ¶ 46, 50.
[266] *Ibid*, ¶ 49.
[267] *Ibid*, ¶ 68.
[268] *Ibid*, ¶ 49, 69.
[269] *Ibid*, ¶ 70.
[270] *Ibid*, ¶ 89.

criminalizing speech without incitement to violence to be permissible under Article 10, in the contexts of obscenity and blasphemy.[271]

(ii) (National Analogues) Recruitment Criminalized

In the absence of additional cases that analyzed freedom of expression with respect to recruitment for VEOs, statutory analysis cases may lend insight into broader legal trends. Two cases from Kazakhstan were identified that convicted individuals for using digital speech for the purposes of terrorist recruitment.

In the first case in Kazakhstan, *The Case of Kairat Bektenov and Others*,[272] the creators of a WhatsApp chat group that allegedly shared messages urging members to join ISIL and participate in jihad were convicted under the Criminal Code of the Republic of Kazakhstan, Articles 233 and 164, which respectively restricts the promotion of terrorist or extremist propaganda and forbids the incitement of social, national, racial, or religious enmity.[273] They each received sentences of five to six years in prison.

In the second case in Kazakhstan, *The State v. Bulat Zhakpbaevich Satkangulov*,[274] Bulat Zhakpbaevich Satkangulov, disagreed with his friends regarding their criticism of ISIS, and defended the extremist group by trying to justify their actions, and had extremist material on his personal laptop.[275] Satkangulov was charged under the Criminal Code of the Republic of Kazakhstan, Article 256, which criminalizes propaganda of terrorism and the storage of materials with the purpose of dissemination.[276] The court ruled that Satkangulov's statements constituted propaganda of terrorism, where propaganda was defined as the "public dissemination of various

[271] *Ibid*, ¶ 89. *See Hoare v United Kingdom* [1997] EHRLR 678 (obscenity); *Wingrove v United Kingdom* [1997] 24 EHRR 1 (blasphemy).
[272] *The Case of Kairat Bektenov and Others*, 1-432/2015 (Kazakh First Instance Court, Aug 3, 2015). https://globalfreedomofexpression.columbia.edu/cases/the-case-of-whatsapp-extremists-kazakhstan/.
[273] *Ibid.*
[274] *The State v. Bulat Zhakpbaevich Satkangulov*, 1-406/2015 (Kazakh First Instance Court, Nov. 18, 2015). Available at: http://causa.kz/2015/11/prigovor-sud-2-goroda-kostanaya-ot-18-noyabrya-2015-goda-1-406-2015-v-otnoshenii-satkangulova-b-zh/.
[275] "The State v. Bulat Zhakpbaevich Satkangulov," Columbia Global Freedom of Expression Database, Oct. 16, 2014.
[276] *Ibid.*

perspectives, ideas, knowledge, and teachings."[277] Accordingly, Satkangulov was convicted of violating Article 256, and sentenced to six years in prison.[278]

3. Operational Support

a. Collecting and Transferring Funds

Only one case surfaced that focused on assessing freedom of expression where parties collected and transferred funds to extremist organizations. Other cases addressed offenses of funding or transferring of funds to VEOs as issues of statutory interpretations, also detailed below.

(i) Offline Fundraising: Ties to Hamas – Deportation Violated Free Expression

Case: Mahajna v. Home Secretary, British Immigration and Asylum Chamber, 2012.

Synopsis: A British appellate court held that deporting an Arab Israeli citizen on grounds of promoting hatred or terrorism would disproportionately interfere with his rights to freedom of expression and association under the ECHR, even though Israel had previously convicted him for funding terrorism.[279] This case, *Mahajna v. Home Secretary*, is more fully analyzed in the section on Ideological Support: Encouraging Future Attacks, but the part relevant to funding is discussed here.

In 2003, Israel convicted Mahajna of using charitable organizations as a front to fund Hamas, but Mahajna maintained he had only provided funds for humanitarian purposes. The British appellate court found the conviction irrelevant in terms of bringing Mahajna within the scope of the Unacceptable Behaviors list, reasoning that there is no evidence that funds provided by Mahajna were put to other than charitable purposes, or benefited the military wing of Hamas (illegal in the UK) rather than the legal political wing of Hamas (legal in the UK).[280] The court found that any support for terrorism commited thereby was "purely formal," and that there was no evidence Mahajna's funds posed a danger to Israel or have led to any community hatred or

[277] *Ibid.*
[278] *Ibid.*
[279] *Mahajna v. Home Secretary*, UKUT B1 (IAC) (U.K. Upper Tribunal, Immigration and Asylum Chamber 2012).
[280] *Ibid*, ¶ 74.

violence.[281] As a result, Mahajna's conviction for terrorist financing did not justify restricting his free expression by deporting him.

(ii) (National and Extradition Analogues) Funding Criminalized

Other cases involved transferring funds to unequivocal terrorist organizations like the Islamic State, but these courts did not analyze the underlying free expression concerns potentially raised by criminalizing this conduct. For example, in *Imran Kassim v. Singapore*, Kassim was found to have violated Singapore's Terrorism (Suppression of Financing) Act by sending 450 Singaporean dollars for the purposes of publication of ISIS propaganda, with his claim to freedom of expression being the rejection of Singaporean national law.[282]

The *Erdoğan v. Morocco* case serves as another example of criminalization of the collection and transferring of funds to allegedly terrorist or extremist organizations.[283] Citing a supposed money-laundering charge for previously helping the Turkish Islamist organization Hizmet raise money, Turkey charged Erdoğan with financing terrorism and requested his extradition from Morocco; Erdoğan denied the charges.[284]

The Committee Against Torture, which heard Erdoğan's international appeal, ruled that his extradition to Turkey would violate the Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment's (CAT) Article 3.[285] The Committee cited the persecution of individuals with backgrounds similar to Erdoğan and their subjection to torture in Turkey, as well as the general deprivation of human rights there, including deprivation of freedom of association and expression. In particular, the committee found that as someone with perceived or

[281] *Ibid*, ¶ 75.
[282] Amalina Abdul Nasir, "CO20017: The Case of Imran Kassim: What Does It Tell Us?" S. Rajaratnam School of International Studies, Jan. 24, 2020. Available at: https://www.rsis.edu.sg/rsis-publication/icpvtr/the-case-of-imran-kassim-what-does-it-tell-us/#.Xn5KZIhKhPY.
[283] *Ferhat Erdoğan v. Morocco*, No. 827/2017 (Committee Against Torture, May 10, 2019). Available at: https://undocs.org/en/CAT/C/66/D/827/2017.
[284] *Ibid.,* ¶ 2.1-2.7.
[285] *Ibid.,* ¶ 9.11.

actual membership of Hizmet, Erdoğan would be at particular risk of torture.[286] Accordingly, Erdoğan's extradition to Turkey would constitute a violation of Article 3 of CAT.[287]

b. Coordination and Planning

Two cases surfaced that analyzed convictions for coordinating and planning for terrorist and extremist organizations in the context of freedom of expression. In the first case, detailed further below, an Ethiopian regional court found the evidence presented by government did little to show a meaningful connection to terrorism, ruling that the convictions violated the bloggers' freedom of expression.[288]

In the second case, the UN Human Rights Committee found a South Korean man's conviction for disseminating pamphlets violated free expression, holding that his simultaneous organization of violent protests did not justify separately punishing him for the pamphlet dissemination.[289]

Other cases did not mention free expression but upheld convictions for preparations on bases of statutory analysis. For example, in *Regina v. Rashid*, a U.K. appellate court found that Husnain Rashid's distribution of online extremist material merited punishment under Section 1 (encouragement of terrorism) and Section 5 (preparation of terrorist acts) of the Terrorism Act of 206.[290] Rashid was convicted of assisting others to commit acts of terrorism for conduct that included: creating and administering a network of Telegram channels aimed at inspiring lone wolf terrorists; sending links to and uploading videos, documents, and materials detailing how to manufacture explosives and shoot down aircraft; and analyzing previous attacks and researching ways to carry them out more successfully in the UK. The appellate court upheld sentences of life imprisonment, with minimum terms of nineteen years.

286 *Ibid.,* ¶ 9.6-9.11.
287 *Ibid*., ¶ 10.
288 *Federal Prosecutor v. Soleyana Shimeles Gebremariam and Others*, F/M/A/05/07 (Ethiopean First Instance Court, Oct. 16, 2015). Available at: https://www.cardeth.org/wp-content/uploads/2015/02/Final-charge-on-Zone9-bloggers-and-journalists.pdf.
289 *Keun-Tae Kim v. Republic of Korea*, No. 574/1994 (UN Human Rights Committee, Nov. 3, 1998).
290 *R v. Rashid*, 2018/03568/A3 (EWCA Crim 797, 2019). Available at: https://www.bailii.org/ew/cases/EWCA/Crim/2019/797.html.

(i) Blogging: Sharing Social and Political Issues - Free Expression Violated.[291]

Case: Federal Prosecutor v. Soleyana Shimeles Gebremariam and Others, Ethiopia Federal High Court, 2015.

Synopsis: In 2015, an Ethiopian court acquitted a group of bloggers, finding the government's evidence too weak to prove the charges of planning an act of terrorism and finding the group's posts consistent with free expression.

In April of 2014, the Ethiopian government charged ten members (with one charged *in absentia*) of the blog Zone 9 of planning to execute terrorism under Articles 32 § 1a,b and 38 § 1,2 of the Ethiopian Criminal Code and Articles 3 § 2 and 4 of the Ethiopian Anti-Terrorism Proclamation. Zone 9 is centered around identifying social and political issues in Ethiopia.[292] In the charges against the ten members, the Federal Prosecutor alleged that they made short- and long-term plans, divided up responsibilities, and received trainings for skills including covert communication, protest leadership, and explosives handling, as part of an effort to overthrow the constitutional order via organized violence and terrorism.[293]

However, in July 2015, the government released and dropped charges against five of the detained members. On October 16 of the same year, the Federal High Court of Ethiopia for the 19th Criminal Bench acquitted the remaining members of Zone 9 that were still charged, dismissing the evidence provided by the government as too weak to prove they were planning acts of terrorism and finding their blog posts consistent with freedom of expression.[294]

Note: The original decision could not be located online, so more detail about the court's reasoning is unavailable.[295]

(ii) Leaflets: Echoing Group's Nonviolent Talking Points – Free Expression Violated

Case: Keun-Tae Kim v. Republic of Korea, UN Human Rights Committee, 1998.

[291] *Federal Prosecutor v. Soleyana Shimeles Gebremariam and Others*, F/M/A/05/07 (Ethiopia First Instance Court, Oct. 16, 2015).
[292] "Federal Prosecutor v. Soleyana Shimeles Gebremariam and Others (Zone 9 Bloggers)," Columbia Global Freedom of Expression Database, Nov. 27, 2014.
[293] *Federal Prosecutor v. Soleyana Shimeles Gebremariam and Others*.
[294] "Federal Prosecutor v. Soleyana," Columbia Global Freedom of Expression Database.
[295] *Ibid*.

Synopsis: The UN Human Rights Committee found a South Korean man's conviction for disseminating pamphlets violated free expression, even though he had also been convicted of organizing violent demonstrations around the same time.

In 1998, the UN Human Rights Committee (HRC) held a South Korean conviction of a political party official for distributing material for the benefit of the North Korean regime violated free expression under Article 19 of the ICCPR.[296] For events occurring soon after the pamphlet distribution, Kim was also convicted for organizing illegal demonstrations in which participants threw Molotov cocktails, set cars afire, and injured over a hundred policemen.[297]

The UN Human Rights Committee (HRC) held that Kim's organization of violent protests did not justify separately punishing him for disseminating the pamphlets under Article 19 of the ICCPR.[298] However, a dissenting member disagreed, arguing that Kim's pamphlet dissemination was what led to the violent protests in question, so the restriction on Kim's free expression was legitimate under national security and public order.[299] This case is covered more fully in the section on Ideological Support: Praise of Ideology.

4. Ideological Support

a. Sharing Propaganda (Without Sentiment)

Five cases surfaced analyzing convictions for sharing propaganda without additional sentiment under the framework of freedom of expression; four from the European Court of Human Rights (ECtHR), and one from an appellate court in the U.K. All of the cases examined in this section were judged to be justified infringements on freedom of expression.

Overall, issues of intent and proximity to the reposted content were notable in these judgments. For example, a U.K. court relied on recklessness and contextual information to establish intent.[300] In another case against a broadcasting company, the ECtHR cited the fact that TV hosts did "nothing to distance themselves" from their PKK guests' message, as evidence of

296 *Keun-Tae Kim v. Republic of Korea*, No. 574/1994 (UN Human Rights Committee, Nov. 3, 1998).
297 *Ibid.*, ¶ 4.2.
298 *Ibid.*, ¶ 12.5.
299 *Ibid.*
300 *R. v. Iqbal,* 2014/01692 C5 (EWCA Crim 2650, 2014).

intent to spread that message.[301] Further, the reach of the information or message was another measure that the courts often considered when assessing free expression and sharing propaganda.

The sentencing for these cases varied, though imprisonment for sharing or reposting propaganda was highly disfavored. Only one case approved a conviction involving imprisonment: the U.K. court upheld a conviction with a three years' prison sentence. In this small set of cases, individual users more often faced imprisonment, while platforms were more likely to be penalized with fines.

(i) Facebook: Sharing of Propaganda – Conviction does not Violate Free Expression

Case: R. v. Iqbal, England and Wales Court of Appeals, 2014.[302]

Synopsis: A U.K. appellate court upheld a conviction for disseminating terrorist publications. First, the court held that recklessness, as the *mens rea* required by the underlying law with respect to incitment of terrorism, was sufficient to comply with ECHR Article 10 on free expression; intentionality was not required. Second, in finding recklessness established for this conviction, the court looked to contextual factors like the number of Facebook profiles the defendant created to circulate links to the publications as well as the usernames he chose.

The defendant, Khuram Shazad Iqbal, was convicted by British courts for dissemination of terrorist publications,[303] and sentenced to three years in prison.[304] Between January and October of 2013, Iqbal circulated many terrorist publications by posting them publicly on several Facebook profiles. These profiles were titled "Abu Irhaab.948," "Abu Irhaab.52," "Abu Irhaab.351," "Abu Irhaab.3990," "Abu Irhaab.90" and "Abu Irhaab.3576."[305] In Arabic, "Abu Irhaab" means "Father of Terrorism." Per the relevant British law, an important element of the offense is *intentionality* or *recklessness,* which is necessary for the act to amount to direct or indirect encouragement or other inducement to the commission, preparation or instigation of acts of terrorism.

[301] *Roj TV A/S v. Denmark*, No. 24683/14, (ECtHR, Apr. 17, 2008).

[302] *R. v. Iqbal,* 2014/01692 C5 (EWCA Crim 2650, 2014).
[303] *Ibid*, ¶ 2. Terrorism Act 2006, Section 2(1)(a) and (2)(c)(d).
[304] "'Father of terrorism' Khuram Iqbal jailed at Woolwich Crown Court," *BBC News*, Sept. 11, 2014. Available at: https://www.bbc.com/news/uk-wales-29167225.
[305] *R. v. Iqbal*, ¶ 5.

Two key issues are relevant in the analysis of this case: first, whether criminalizing reckless or unintentional dissemination is in conflict with Article 10 of the ECHR; and second, the use of contextual information to establish intent.

First, the UK Appellate Court judge in this case weighed the possibility that section 2 of the Terrorism Act 2006 was not compliant with Article 10 of the ECHR in that it required intent or merely recklessness in establishing that Iqbal's conduct would amount to encouragement or incitement to acts of terrorism. The court concluded that section 2 was a necessary and proportionate response to the threat from terrorism, and, looking to prior rulings, found no need to reject recklessness as a subjective standard of liability.[306]

Second, the court looked to factors apart from explicit "speech" to establish intent and/or recklessness, namely the usernames and practices of the profiles that Iqbal used to repost and share material. These profiles were named for variations of the name "Abu Irhaab," meaning "father of terrorism" in Arabic.[307] Additionally, the court cited the fact that Iqbal generated multiple accounts in order to repost the content as reason to establish intent.

(ii) TV Channel: Broadcasting Statements by the PKK – Conviction does not Violate Free Expression

Case: Roj TV A/S v. Denmark, ECtHR, 2018.[308]

Synopsis: In 2018, the ECtHR found a Danish conviction of a TV broadcast company for promoting the PKK's "terror operation," consistent with free expression, reasoning that the company's broadcasts had incited violence and support for terrorism by repteatedly featuring PKK fighters as heroes, showing PKK leaders inciting violence without pushing back against their message, and widely disseminating this message to a large audience. Instead of finding a restriction of free expression warranted as necessary and proportionate under the classic three-part test set forth in ECHR Article 10, the ECHR based its ruling in ECHR Article 17, "Prohibition on abuse

306 *Ibid*, ¶ 22.
307 "'Father of Terrorism' Facing Jail," *BBC News*.
308 *Roj TV A/S v. Denmark*, No. 24683/14, (ECtHR, Apr. 17, 2008).

of rights;" the ECtHR found the broadcasts ineligible for the protection of free expression because they aimed to destroy rights of others guaranteed in the ECHR.

In 2012, Danish courts convicted a TV broadcast company for promotion of terrorist activities (Danish Penal Code, Article 114e). Upholding fines levied against the company and revoking its broadcasting license, the Danish Supreme Court found their Kurdish language TV programs had promoted the PKK's "terror operation," and so were not protected free expression under ECHR Article 10. In finding the company's TV programs unprotected by freedom of expression, the ECtHR accepted the Danish court's finding that the company's broadcasts portrayed deceased PKK fighters as heroes and could not be dismissed as a mere declaration of sympathy.[309]

Under Article 17 of the ECHR, the ECtHR held the company could not use Article 10's protections for free expression to inoculate its incitement of violence and support for terrorist activity, because its broadcasts were contrary to the fundamental values of the Convention.[310] Article 17 of the ECHR, "Prohibition on the abuse of rights," declares that the ECHR does not protect any entity's attempt to subvert the ECHR, such as by trying to destroy the rights of others that are guaranteed by the ECHR.[311]

The court found the company's broadcasts contrary to the fundamental values of the ECHR because they repeatedly amplified the PKK's inciting message by sharing it with a wide audience, without adding comments disclaiming support for the PKK or the message itself. The court found the company had incited violence and support for terrorist activity by repeatedly featuring PKK leaders, who incited viewers to join the PKK's fight, while the company's TV hosts listened passively and made no attempt to distance themselves from the message such as by asking critical questions.[312] The court's reasoning also incorporated two contextual factors, reach and speaker's

309 *Ibid*, ¶ 6.
310 *Ibid*, ¶ 47.
311 The full text of the article reads: "Nothing in this Convention may be interpreted as implying for any State, group or person any right to engage in any activity or perform any act aimed at the destruction of any of the rights and freedoms set forth herein or at their limitation to a greater extent than is provided for in the Convention."
312 *Ibid*, ¶ 9.

personal history: television broadcasting reached a wide audience, and the TV company had previously been financed by the PKK.

(iii) Newspaper: Publishing Declarations by Individuals – Prosecution as Justified Interference

Case: Case of Saygili and Falakaoglu v. Turkey, ECtHR, 2009.

Synopsis: In 2009, the ECtHR found consistent with free expression Turkey's conviction of two newspapermen for publishing detainees' declarations of a hunger strike to protest aspects of the Turkish prison system. The sentences did not include imprisonment. In finding the newspaper publishers vicariously liable for the detainee's statements, the court cited their special duties, as members of the press, to contribute to democracy by refusing to give the detainees an outlet to incite hatred. In finding the declarations to constitute incitement, the court emphasized the declarations' call for readers to support their resistance and fight against the prison system, and used violence occurring two months after the declarations as evidence that the situation was volatile when the declarations were published.

In 2002, Mr Fevzi Saygılı and Mr Bülent Falakaoğlu published in their newspaper, called *Yeni Evrensel,* three declarations from detainees associated with a proscribed organization; the declarations announced that they would go on a hunger strike and die rather than "enter[ing] the cells" until the Government abolished the F-type prison system.[313] One such declaration added a request to abolish the Anti-Terrorism Law, and proclaimed: "We call on the working class, labourers, all oppressed people and revolutionary people and ask them to support us in this resistance, for which we are ready to sacrifice our lives . . . We call on you to fight off the cell system." In response, the Chief Public Prosecutor at the State Security Court filed an indictment charging the applicants with publishing the declarations of terrorist organizations. The Istanbul State Security Court found the applicants guilty as charged and levied a fine against them, as well as banning the publication of *Yeni Evrensel* for three days.

[313] *Case of Saygili and Falakaoglu v. Turkey* (No. 2), No. 38991/02 (ECtHR Grand Chamber, Feb. 27, 2009).

The ECtHR held that this interference with freedom of expression was necessary and justified, due to the newspapermen's decision to publish these declarations during a sensitive period. The court rejected the arguments from Saygılı and Falakaoğlu that in publishing these declarations, they were merely imparting information as a commercial endeavor. As members of the press, the court found the two had "duties and responsibilities" to contribute to "the proper functioning of political democracy," which were heightened during periods of conflict, and which they had violated by publishing these declarations. The ECtHR thus held them vicariously responsible for the declaration's incitement because they provided the declaration's writers, who were associated with illegal armed groups, "with an outlet to stir up violence and hatred."

The ECtHR found the declarations constituted incitement because their content portrayed violence as "a necessary and justified measure of self-defence" in response to imprisonment, and urged readers to take action to support the struggle. The court also took note of contextual factors, pointing out that less than two months after the publication, unrest broke out in the country's prisons. The court did not claim that the declarations resulted in this unrest, but considered the unrest as evidence that the state had good reason to fear violent reactions to this type of declaration during this sensitive period. The court also found that the heavy punitive fines, equal to ninety percent of the newspaper's average monthly sales, were a proportionate restriction.

(iv) Book: Depicting Members as Heroes – Prosecution as Justified Interference

Case: Fatih Taş v. Turkey, ECtHR, 2003.

Synopsis: In this case, the ECtHR held that publishing others' incitement can be criminalized under Article 10, upholding a Turkish conviction of a book's editor and publisher for disseminating terrorist propaganda.

The 2003 ECtHR case of *Fatih Taş v. Turkey* is relevant here as well as in the section on Praising Ideology, where it is discussed in more detail. However, the key relevant points to this function are outlined here. Fatih Taş published a book containing three passages depicting PKK members, and Turkish authorities charged him with disseminating terrorist propaganda.[314] By the

[314] *See The Prevention of Terrorism Act* (Law no. 3713), § 7(2).

time the case reached the ECtHR, his conviction and sentence had been quashed after seven years of criminal proceedings. Fatih Taş did not write the passages in question, but nonetheless the ECtHR held that the interference with his freedom of expression was legitimate and proportionate because the passages were reasonably considered to incite violence, and the defendant exercised control over them as publisher and editor.[315]

(v) Magazines, Books: Bringing in Propaganda for Distribution – Free Expression Not Violated

Case: Kaptan v. Switzerland, ECtHR, 2001.

Synopsis: In 2001, the ECtHR held that Switzerland's confiscation and destruction of a large quantity of propaganda in support of the PKK was a legitimate interference with its owner's freedom of expression. The court considered (1) the large number of publications as well as their content, which encouraged armed struggle against Turkey, as well as (2) the geopolitical and domestic circumstances that contributed to the weight of the publication's message.

In September 1997, Faruk Kaptan attempted to transport about 88 kilos of physical propaganda related to, and in support of, the Kurdish Workers Party (PKK) into Switzerland.[316] Swiss customs authorities, when coming across the large amount of propaganda and political materials, sent them to the Swiss Federal Attorney's Office, which concluded that these materials propagated violence as "the only alternative against a 'Turkish Terror State,'" and seized and destroyed them. Kaptan applied to the ECtHR for relief on the grounds that the confiscation and destruction of the materials was an unnecessary, disproportionate and arbitrary interference with his rights under Article 10 of the ECHR, protecting his right to freedom of expression.

In ruling that Switzerland's interference with the applicant's free expression was justified under Article 10 of the ECHR, the ECtHR considered (1) the large number of publications as well as their content, which encouraged armed struggle against Turkey, and (2) the geopolitical and domestic circumstances that contributed to the weight of the publication's message.

315 *Fatih Tas v. Turkey* (No. 3), No. 45281/08 (ECtHR Grand Chamber, Apr. 24, 2018).
316 *Kaptan v. Switzerland*, No. 55641/00 (ECtHR Second Section, Apr. 12, 2001)

First, the court concluded that because the individual, Kaptan, was transporting nearly 88 kilos of publications, they were intended for sale and distribution within Switzerland, rather than personal use. Second, the materials advocated for violence, recommending armed struggle against Turkish authorities.[317] Third, the writings seemed to be aimed at a Kurdish emigrant audience, attempting to radicalize the Kurdish living within Switzerland, transmit the tensions that exist within Turkey to Switzerland, and exert pressure on the Swiss-Kurdish communities.

The court concluded that the incitement to violence extended throughout the publications and rather than being in individual passages that could be omitted. For these reasons, the court concluded that this kind of speech, and in this volume, justified restrictions under Article 10 of the ECHR, asserting that the confiscation of these materials was provided by law, necessary, and proportionate.

(vi) (National Analogue) VKontakte: Reposting an Article – Conviction[318]

Case: The Case of Dmitry Semenov, Russian Appellate Court, 2015.

Synopsis: In 2015, a Russian appellate court upheld a journalist's conviction for reposting an article on Russian social media platform VKontakte and sentence of a fine, which the trial court had immediately pardoned. The journalist had posted an article; VKontakte had auto-populated the post with an image of Russia's prime minister and the caption of "Death to Russian vermin." The prosecution successfully argued the journalist had intentionally posted the photo by pointing to his prior actions which it portrayed as extremist: a 2014 repost of a photo including a swastika and his open support of Ukraine on VKontakte.

Dmitry Semenov is a journalist for Open Russia (an independent media outlet), a human rights advocate, and an activist who frequently organized rallies in his native Tsivilsk City. On January 29, 2015, he was charged with incitement of extremism for reposting a photo of Dmitry Medvedev (Russia's Prime Minister) on V-Kontakte (Russia's largest social media network). The photo was a caricature of Medvedev wearing a traditional Caucasus hat with some Arabic text in the background and a statement at the bottom of the photo that read: "Death to Russian vermin."

317 The decision contained no detail or quotes from the materials.
318 *The Case of Dmitry Semenov*, 22-2559/2015 (Russian Appellate Ct. 2015).

Despite Semenov's claims that he did not intend to post the photo, which had auto-populated on his VKontakte re-post of the article, the Federal Committee on National Security alleged that he had intentionally posted the photo. The Committee determined that, because the photo called for the elimination of Russians and degraded persons on the basis of their nationality, the intentional reposting of the photo violated the following: Russian Constitution Articles 19 and 29 which provide for the equality of citizens regardless of nationality; Federal Law 114-FZ on Opposing Extremist Activities which forbids the dissemination of incitement of national hatred over the Internet; and finally—because of the aforementioned violations—the reposted photo constituted public incitement of extremist activities under Russian Criminal Code Article 280, paragraph 1.

The defendant appealed this determination to the High Court of the Republic of Chuvashia where the guilty verdict, resulting in a pardoned fine rather than any imprisonment, was confirmed. The central issue in the case was the question of whether the defendant intentionally reposted the extremist photo. To be found guilty of disseminating extremist materials under the Russian Criminal Code Article 280 Paragraph 1, the court needed proof of the existence of intent to post. Despite the defendant's arguments that he did not intentionally mean to share the photo with the reposted article, the committee used the defendant's past actions to infer intent. The prosecution had portrayed him as an extremist by pointing to his open support of Ukraine on social media and his 2014 administrative conviction for reposting a photo of sport fans waving a fan with a swastika.

Note: No additional detail is available as to the court's reasoning is available because the original court documents could not be found, let alone in English.

b. Encouraging future attacks

Seven cases surfaced analyzing convictions for encouraging future attacks under the framework of international rights to freedom of expression; four from the European Court of Human Rights (ECtHR), one from the Inter-American Court of Human Rights (IACHR), and one from a British court. Five of the cases examined in this section were judged to be unlawful infringements on free expression, while one was deemed justified interventions. Three additional

national court cases analyzed national rights to freedom of expression as applied to materials encouraging future attacks, in Australia, Russia, and Canada.

Notably, some judgements emphasized the justification used to restrict the speech. For instance, in the IACHR case involving Venezuela, the court held that the infringement on free expression was not permissible under the justification that the speech "incited a coup," but may have been justifiably restricted if the state had utilized an "incitement to violence" or disorder exception.

While these cases span a large breadth of situations, many focused on whether the piece of expression itself *on the whole* is intended to incite an audience: that is, if the entire article, speech, or post constituted incitement, rather than if an errant remark or sentence inside the piece could be characterized as incitement.

Many also focused on the proportionality of punishment. As with many other functions, imprisonment was disfavored. The ECtHR held that three sentences of imprisonment violated freedom of expression: one for 1 year from a Turkish court, one for 5 years from a Russian court, and one for 11 years from an Azerbaijani court. Of the two criminal convictions upheld as justified infringements on free expression, one involved merely a fine while the other had a suspended sentence of one year in prison.

(i) TV Station: Content Encouraging a Coup – Conviction Violates Free Expression

Case: Granier et al. v. Venezuela, Inter-Am. Comm'n on H.R., 2012.

Synopsis: In 2007, the Venezuelan government shut down a television station, alleging that it had incited a coup in 2002. The IACHR held this a violation of free expression under the ACHR, reasoning that the state authorities had made no attempt to follow due process of law or prove any incitement of violence or other illegality by the television station.

In 2002 and years after, the brief coup d'etat against Chavez's government was fresh in the minds of nearly all Venezuelans. Radio Caracas Television (RCTV), is a television station that was reporting on Venezuelan political events at this time. As the coup unfolded, RCTV was reporting on it and criticizing President Chavez, while also giving a platform to the participants of the coup with interviews and profiles.

RCTV attempted to renew its broadcasting license in 2007, but the government denied its application, claiming that the station had incited the 2002 coup.[319] The State shut down RCTV, seizing the television station's equipment and occupying its studios, and established a State-sponsored television channel on RCTV's airwaves. RCTV was unable to obtain redress through the State's domestic courts, and so appealed the actions to the Inter-American Court on Human Rights.

The IACHR determined that these penalties imposed on RCTV constituted an indirect curtailment of RCTV's freedom of expression, which violated Article 13 of the ACHR. The court explained that the restriction might have been permissible had Venezuela established that RCTV incited violence or otherwise violated the law in a proceeding that complied with due process, yet Venezuela had made no attempt to do so. Thus, this outcome case came down to the difference in charges: Venezuela had not established any incitement of violence to justify the restriction.

(ii) Online Forum Post: Hypothetical Attack Plan – No Violation of Free Expression

Case: Smajić v. Bosnia and Herzegovina, ECtHR, 2018.

Synopsis: In 2018, the ECTHR found consistent with free expression under Article 10 a Bosnian conviction of a Bosniac lawyer for inciting hatred and discord because he posted a hypothetical attack plan against Serbians online. The court described the online forum, even though it required user registration, as "publicly accessible," and emphasized the lightness of the punishment: a suspended sentence of imprisonment for one year.

In 2010, Bosnian lawyer Mr. Abedin Smajić posted a hypothetical plan of attack against an ethnic minority in an online forum that required users to register in order to view content.[320] He wrote that—if a sub-region of Bosnia seceded—Bosniacs "should organize ourselves" and attack three Serbian villages and then attack two cities, "cleans[ing]" the associated city centers.[321] In 2012, for inciting hatred and discord, a Bosnian court sentenced him to one year of imprisonment, but suspended for three years.

[319] *Granier et al. v. Venezeula*, Report on Merits, Report No. 112/12, Case No. 12.828 (Inter-Am. Comm'n H.R., Nov. 9, 2012).
[320] *Smajić v. Bosnia and Herzegovina*, No. 48657/2016 (ECtHR 2018).
[321] *Ibid*.

The ECtHR found no violation of Article 10, but analyzed the case under the framework of incitement of racial hatred rather than violence. However, relevant analogies for incitement to violence include that while the website in question required users to register, the court characterized the forum as "publicly accessible."[322] Even though the attack plan was phrased hypothetically, the court deferred to the domestic courts' interpretation of the facts and emphasized the lightness of the punishment.

(iii) Published Article: Emailed Hypothetical Attack Plan – Conviction Violates Free Expression

Case: Fatullayev v. Azerbaijan, ECtHR, 2010.

Synopsis: In 2010, the ECtHR held that Azerbaijani courts had violated free expression under ECHR Article 10 by convicting a newspaper publisher of terroristic threats. The defendant had published an article discussing a hypothetical scenario of an attack by Iran on Azerbaijan. The ECtHR emphasized that the article did not encourage the attack scenario he described, and considered that the article's author was a private citizen in no position to influence Iranian attack decisions.

Eynulla Emin oglu Fatullayev was the founder and chief editor of the newspapers *Gündəlik Azərbaycan*, published in the Azerbaijani language, and *Realny Azerbaijan*, published in the Russian language.[323] In 2007, *Realny Azerbaijan* published an article under a pseudonym which discussed a hypothetical scenario in which Iran would bomb sites in Azerbaijan, and included a list of targets. Later that year, Azerbaijani authorities charged the publisher with the criminal offense of threat to terrorism; the trial court found him guilty on all charges and convicted him of threat of terrorism (eight years' imprisonment), incitement to ethnic hostility (three years' imprisonment) and tax evasion (four months' imprisonment).

The ECtHR ruled Azerbaijani courts violated Article 10 by convicting a journalist of a terrorist threat on the basis of his analytical article criticizing Azerbaijan's domestic and foreign

[322] *Ibid*, ¶ 36.
[323] *Fatullayev v. Azerbaijan,* No. 40984/07 (ECtHR 2010).

policy. The article discussed a hypothetical scenario, in which Iran would bomb sites in Azerbaijan, and included a list of targets. The domestic courts had found this article, emailed to several people who testified that it disturbed them, to be aimed at "frightening the population," and therefore to convey a terrorist threat.

The ECtHR dismissed this finding as arbitrary; the author was a private citizen "not in a position to influence" any of the events his article speculated about, much less control Iranian decisions to attack Azerbaijani sites. Furthermore, the article did not approve of or argue for any such hypothetical attacks.

(iv) Public Speaking: Implicit Incitement – Conviction Violated Free Expression

Case: Han v. Turkey, ECtHR, 2005.

Synopsis: The ECtHR held that a conviction for disseminating separatist propaganda violated Article 10 because the speaker's statement did not itself encourage violence. The court considered whether or not the content encouraged violence to be the "essential" factor to establish necessity in a democratic society.[324]

In 1994, Mr. Tahir Han was a member of the Peoples' Democracy Party (*Halkın Demokrasi Partisi*, "HADEP") and made a speech during HADEP's first annual congress. During the speech, he allegedly incited illegal activity by saying: "As we have said before, it is impossible not to collide with all the legal constraints and formalities imposed by the Republic of Turkey. . . Therefore, HADEP should . . . direct the peoples' anger at heightened resistance. A party programme . . . which is confined to legal boundaries is bound to be unsuccessful."[325]

As a result of this speech and some political tensions within HADEP that arose after, the Ankara State Security Court convicted Han of distributing propaganda "against the indivisible integrity of the State, " and sentenced him to one year's imprisonment and a fine.[326] It based this conclusion on the applicant's having justified and implicitly called for illegal activity in order to advance the cause of Kurdish nationalism. Therefore, the case did not involve incitement of

[324] *Han v. Turkey*, No. 50997/99 (ECtHR 2005).
[325] *Ibid*, ¶ 7–10
[326] *Turkish Prevention of Terrorism Act* as of 1997, Article 8, § 1.

violence or hate speech; rather, it involved advocacy of separatism that could be considered incitement of law-breaking.

The ECtHR held this conviction violated the applicant's Article 10 rights. While the EtCHR found that the interference was prescribed by law and pursued a legitimate aim ("protecting territorial integrity"), it still concluded that the interference was not justified as necessary in a democratic society, because the speech primarily criticized Turkey's policies on Kurds, and did not incite violence.[327]

Specifically, the Court summarily concluded: "taken as a whole, the applicant's speech does not encourage violence, armed resistance or insurrection." Presumably, by "taken as a whole," the court meant taking the above sentence in the context of the entire speech (which included criticism of the government) and perhaps surrounding circumstances, such as ; However, the court does not explain or define this phrase. Instead, the court called actual incitement of violence "the essential factor" in assessing whether or not the third prong of the classic three-part test: necessity in a democratic society. Accordingly, the ECtHR held that the interference was disproportionate and unnecessary, and violated Article 10. The court did not further explain this part of its reasoning, which comprised only four sentences.

(v) Offline Speeches and Poems: Calling for Martyrdom – Deportation Violated Free Expression

Case: Mahajna v. Home Secretary, British Immigration and Asylum Chamber, 2012.

Synopsis: A British appellate court held that deporting an Arab Israeli citizen on grounds of promoting hatred or terrorism would disproportionately interfere with his rights to freedom of expression and association. The court paid attention to factors such as the speaker's personal history, the lack of demonstrable violence resulting from his words, and the overall meaning and focus of his speeches that included a few sentences glorifying martyrdom and referencing blood libel against Jews. The court also discounted his previous Israeli conviction of funding Hamas

[327] *Ibid*, ¶ 13.

because no evidence contradicted his assertion that his funding had benefited only charitable purposes.[328]

In 2011, the Middle East Monitor invited Arab Israeli citizen Raed Salah Mahajna to come to the United Kingdom and make a speech as well as attend a meeting in the House of Lords. Mahajna had previously been elected mayor of an Arab Israeli town and led the northern arm of the Islamic Movement, outlawed in 2015 by the Israeli Security Cabinet for connections to Hamas and the Muslim Brotherhood. He had previously visited the UK and attended conferences and meetings. In June 2011, the Secretary of State for the Home Department (SSHD entered an exclusion order against Mahajna. Unaware of the order, he entered the country anyway; Mahjanna was arrested and detained awaiting deportation. The Home Secretary had ordered Mahajna's exclusion and deportation based on his having publicly expressed "views that foster hatred which might lead to inter-community violence in the UK," under the scope of the List of Unacceptable Behaviors.[329]

Mahajna challenged his detention on grounds of violation of his Article 10 and Article 11 rights under the ECHR. The first-tier tribunal concluded—in deference to the Home Secretary's wide range of discretion in judging activities possibly dangerous to national security—that Mahajna's deportation was a justified and proportionate interference with his freedom of expression with the aim of preventing disorder.[330] The upper tribunal reversed.

First, the appellate court confirmed that the exclusion could be considered "in accordance with the law" because executed under a published UK policy. Under the UK "Prevent Strategy," anyone who performed activities on the Unacceptable Behaviors List would be presumed excludable; the burden of proof would be on the individual to show he had publicly repudiated his previous extremist views or actions.[331] The appellate court also found the trial court erred in overly deferring to the Home Secretary's interpretation of the facts since they were based partially on mistaken facts, such as an inaccurate translation. The court analyzed *de novo* the five pieces of

[328] This charge is discussed in more detail under Operational Support: Collecting and Transferring Funds. *Mahajna v. Home Secretary*, UKUT B1 (IAC) (U.K. Upper Tribunal, Immigration and Asylum Chamber 2012).
[329] *Ibid*, ¶ 8.
[330] *Ibid*, ¶ 22.
[331] *Ibid*, ¶ 35.

evidence proffered by the Home Office to determine whether or not they constituted Unacceptable Behaviors: activities promoting terroristic violence or hatred that might lead to inter-community violence in the UK.

The first three pieces of evidence did not constitute offenses under the Unacceptable Behaviors policy according to the court: a mistranslated poem, an Israeli indictment of inciting a riot, and an Israeli conviction for funding Hamas (discussed above and in sub-section Operational Support: Collecting and Transferring Funds). In 2003, Mahajna wrote a poem published in his organization's periodical. The Home Office had examined an inaccurate translation of this poem, which called Jews "criminal bombers of mosques," "slaughterers of pregnant women and babies" and "germs," adding: "Victory belongs to Muslims." The actual poem written by Mahajna said these things about "oppressors," without mentioning "Jews" specifically. The appellate court interpreted the poem to address generic oppressors rather than Jews, and found it not to constitute an Unacceptable Behavior. In 2011, a few days before the exclusion order, an Israeli court issued two indictments against Mahajna. The first of these alleged a 2007 speech in which Mahajna alleged an Israeli plan to replace the Al-Aqsa mosque with a temple led to a riot near Al-Aqsa that injured three police officers.[332] The court discredited this claim as unproven in any court, and noted the video of the speech showed no violence.[333]

However, the court considered two of Mahajna's statements to plausibly constitute Unacceptable Behaviors: a reference to blood libel, not discussed here as it relates to hate speech, and a call to martyrdom. In a 2007 speech, Mahajna said: "the most beautiful moments of our destiny will be when we meet Allah as martyrs in the premises of the Al-Aqsa mosque." The court discredited Mahajna's explanation of this statement as merely encouraging passive martyrdom and peaceful protest by focusing on the context of the quote: the sermon began with a threat that God would take retribution against Israel and referred to "blood, killings, and massacres."[334] In contrast, the court declined to weigh two other such statements against Mahajna because they had only

[332] No party argued that the second indictment, about obstructing a police officer by objecting to a search, was relevant to the UK determination. *Ibid*, ¶ 35.
[333] *Ibid*, ¶ 70.
[334] *Ibid*, ¶ 68.

isolated quotes to consider, without being able to consider the full texts or contexts of the entire speeches. These included a 2009 speech in which Mahajna said "if they suggest that we give up our principles and holy sites, we would rather die" and a 2011 statement: "We will not compromise on our principles or holy sites. We prefer to die as shahids and will welcome death joyfully." The court said that without additional context, the court could not responsibly judge whether the references to martyrdom were calls to violence or not.[335]

When weighing the blood libel and calls to martyrdom from 2007, the two actions which the court deemed potentially within the scope of the Unacceptable Behaviors policy, the court assessed both the speaker's history and the lack of violence caused by his statements or previous visits to the country. (78, 80) The court pointed out that these two statements were isolated and dated, both from a single day years ago, and atypical of Mahajna's usual speeches, message, and agenda.[336] The court weighed the slightness of these offenses against the restriction and found it disproportionate and "entirely unnecessary" to ward off violence and instability: "because of a few sentences in the sermon in February 2007, which nobody seems to have regarded as harmful at the time," Mahajna could not be "prevented from being in the United Kingdom or saying anything here (save by telecommunication), for an indefinite period of time."[337] Thus, the court concluded Mahajna's deportation interfered disproportionately with Mahajna's right to freedom of speech under ECHR Article 10 as well as his right to freedom of association under ECHR Article 11.[338]

(vi) Newspaper: Justifications for Violence – Conviction Violated Free Expression

Case: Stomakhin v. Russia, ECtHR, 2017.

Synopsis: In 2017, the ECtHR held Russian courts had violated free expression under ECHR Article 10 by convicting an author and publisher of articles and sentencing him to five years in prison. This was because, even though some of his statements did call for violence, the published articles taken as a whole, on balance, did not demonstrate a pressing need for intervention. The

335 *Ibid*, ¶ 69.
336 *Ibid*, ¶ 78–80.
337 *Ibid*, ¶ 85.
338 *Ibid*, ¶ 85–89.

court emphasized that rather than being considered in isolation, statements calling for violence must be analyzed in context with the full text of a post or article in which they appear.

In 2011, defendant Stomakhin wrote and published newsletter articles about the armed conflict in Chechnya. For instance, one article commenting on a court case said: "Let dozens of Chechen snipers take up their positions in the hills and the city ruins and hundreds and thousands of aggressors perish from their holy bullets! No mercy! Death to the Russian invaders!" Another commented on a hostage-taking by saying: "I, as a national of the Chechen Republic of Ichkeriya (CRI), who is daily suffering from the Russian State Terror, can understand the reasons which pushed Chechen patriots to this extraordinary act."[339]

He was convicted and sentenced to five years in jail by domestic Russian courts because they found these articles to justify him guilty of "having publicly appealed to extremist activities through the mass media" and of having committed "actions aimed at inciting hatred and enmity as well as at humiliating the dignity of an individual or group of individuals on the grounds of ethnicity, origin, attitude towards religion and membership of a social group, through the mass media."[340]

The ECtHR held this violated ECHR Article 10 because, taken as entire texts, the articles did not present a pressing social need to interfere with the applicant's rights. The court acknowledged that some of the article's portions *had* gone beyond the bounds of acceptable criticism and had amounted to calls for violence and justification of terrorism. However, they also found that many other statements in the articles in question had been within the acceptable limits of criticism of governmental actors.

Taking these comments on balance, the court claimed that overall there was no need to so harshly penalize the applicant, and argued that in doing so the domestic court had violated his rights. The ECtHR also added that it was vitally important for States to take a cautious approach when determining the scope of crimes of hate speech. It called on them to strictly construe

339 *Stomakhin v. Russia,* No. 52273/07, (ECtHR 2017). https://hudoc.echr.coe.int/eng#{%22itemid%22:[%22001-182731%22]}
340 Russian Criminal Code, Article 280, §§ 1–2.

legislation in order to avoid excessive interference under the guise of action against such speech, when what was in question was actually criticism of the authorities or their policies.

(vii) (National Analogue) Video: Inciting Hatred & Exacerbating Social Tensions – Banned

Case: The Case of Ayupov R.N., Russian Trial Court, 2016.

Synopsis: In 2016, a Russian trial court banned a video that criticized law enforcement on the theory that it could incite viewers to violence against authorities.

In 2016, the Russian Khanty Surgut City Court ordered that a video containing statements that were critical of law enforcement be added to a unified register of banned extremist material because it allegedly contained extremist content. The video depicted a Russian nationalist leader criticizing Surgut City's law enforcement capabilities. The justification from the Surgut City Court was based on expert findings, which concluded that the video contained statements that could incite hatred toward government officials and exacerbate social tensions.[341]

To reach this decision, the Court examined two central questions: first, if the content was in fact extremist and second, whether it was appropriate to censor it. First, the Court enlisted an expert which concluded that the content was extremist for three reasons: the content could incite hatred of an individual based on their employment, it humiliated the Surgut City police, and could incite viewers to violence against the authorities. Second, after concluding the content was extremist, the Court determined that it would not violate Article 29 of the Russian Constitution, which established the right to free expression. The Court argued that Article 13 of the Federal Law No. 114-FZ criminalized the dissemination of extremist speech and thus applied in this case.

(viii) (National Analogue) Facebook Video Post: Mock Beheading – Conviction

Case: Cottrell v. Ross, Australian Trial Court, 2019.

Synopsis: In 2019, an Australian trial court convicted three nationals of inciting contempt of Muslims for posting a mock beheading video on Facebook and sentenced them to fines.

341 *The Case of Ayupov R.N.*, No 2-1756/2016 (Russian First Instance Court 2016).

In 2015, three individuals staged a "mock beheading" to protest against the building of a mosque in central Victoria, Australia. They filmed this mock beheading and posted it on the far-right nationalist Facebook page, United Patriots Front. The three involved in the video, Cottrell, Shortis, and Erikson, were found guilty of inciting serious contempt of Muslims by the county court of Victoria. The three were each fined two thousand dollars.[342]

The Australian trial court examined whether fining these three men would "impermissibly burden the implied freedom of political communication protected by the Australian Constitution."[343] The court found it did not because 1) it questioned that the video was truly "political" and 2) even if it was political, the ruling presented merely an "incidental" restriction on communication, rather than a "meaningful" one. Further, the court found that the video did not "have a legitimate purpose" in its aims, thus was included in the conviction under Section 25(2) under the Racial and Religious Tolerance Act.

(iix) (National Analogue) Facebook: Incitement to Violence without Intent – Acquitted Case: R v. Othman Hamdan, Canadian Trial Court, 2017.

Synopsis: In 2017, a Canadian trial court acquitted a refugee of inciting a terrorist act via his 95 Facebook posts supportive of ISIS, finding the evidence insufficient to prove beyond a reasonable doubt that he intended to incite readers to violence.

In 2017, Palestinian refugee Othman Hamdan was charged by Canadian authorities after posting language discussing Middle East politics, and supporting ISIS in Iraq and Syria on various Facebook accounts. The Canadian trial court, the Supreme Court of British Columbia, examined 85 posts from Facebook accounts, looking for two elements to charge Hamdan with a criminal violation of Canada's anti-terrorism laws: 1) the posts were likely to incite a reader to commit a terrorist act and 2) that Hamdan intended to incite his audience.[344]

[342] *Cottrell v. Ross*, AP-17-2306, (County Court of Victoria, 2019). Available at: https://www.countycourt.vic.gov.au/files/documents/2019-12/cottrell-v-ross-2019-vcc-2142.pdf
[343] *Ibid*, ¶ 125
[344] *R v. Othman Hamdan*, 2017 BCSC 1770 (Supreme Court of British Columbia 2017). Available at: https://www.bccourts.ca/jdb-txt/sc/17/17/2017BCSC1770.htm

The Canadian trial court determined that a reasonable person would have found at least one of the 85 posts to be an active inducement to commit a terrorist act. This post included descriptions of how "lone wolf" actors can carry out jihad and encourages them to do so.[345] However, the court also determined that they could not prove that Hamdan intended to incite a reader to violence in this post beyond a reasonable doubt. Accordingly, the Court ultimately acquitted Hamdan.

c. Praising past attacks

Four cases surfaced analyzing convictions for praising or defending past terrorist attacks under the framework of freedom of expression; two from a Spanish appellate court, one from a Norwegian appellate court, and one from the European Court of Human Rights (ECtHR). Two cases held the convictions justified,[346] while the other two judged them violations of freedom of expression.[347] Three of the four cases involved satire.

The two cases upholding convictions both involved provocative content with ambiguous relationships to violence published or posted in the immediate aftermath of major terrorist attacks. The first involved a satirical cartoon of the Twin Towers saying "We all dreamt of it . . . Hamas did it,"[348] and the second was a satirical Tweet that appeared to evince identification with a victim of one attack and a perpetrator of another: "[T]onight as far as I'm concerned I feel like Charlie Coulibaly."[349] Given the ambiguity of this content, these two courts weighed context more heavily when upholding convictions, focusing on the timeliness of the posts, the sensitivity of the political situations, and the ability of the posts to reach mass audiences online.

In contrast, the two courts overturning convictions both concluded convictions would violate freedom of expression because the content in question had no clear link to violence: it neither incited future violence nor even praised or defended the justifiability of past violence. In

345 *Ibid*, ¶ 19.
346 *Leroy v. France*, No. 36109/03 (ECtHR, Feb. 2, 2008); *Tribunal de grande instance de Paris* (jugement correctionnel du 18 mars 2015) (High Court of Paris, Mar. 18, 2015).
347 *The State v. Cassandra Vera,* STC 493/2018, (Spanish Supreme Court, Feb. 26, 2018); *Hussain v. The Norwegian Prosecution Authority*, Nos. 14-0499903AST-BORG/01, 14-174730AST-BORG/01 (Norwegian Appellate Court, June 22, 2015).
348 *Leroy v. France*.
349 *Tribunal de grande instance de Paris* (jugement correctionnel du 18 mars 2015).

the first case, involving a teenager's jokes on Twitter about a decades-old attack, the content as well as the context (long ago attack, teenage speaker) weighed against restricting freedom of expression.[350] In the second case, however, the context would have weighed in favor of restricting free expression; the attacks referenced were recent and the poster had served as a spokesman for an Islamist extremist group. But because the content's relationship with violence was ambiguous rather than explicit (eulogizing attackers with comments like "May Allah [] reward them"), the conviction was overturned.[351]

In this set of cases, imprisonment for praising past attacks was highly disfavored. In the only one of the four cases where the defendant actually served time in prison (a one-year sentence), the conviction was held to violate freedom of expression.[352] In contrast, a conviction resulting in a fine and a conviction meriting a suspended sentence of two months were both upheld.[353]

This is not to imply that no countries are imprisoning people for solely praising former attacks, only that the on-point appellate cases we found which analyzed freedom of expression concerns did not endorse imprisonment as a response. For instance, UK courts sentenced a man to three years for posting terrorist propaganda, including content praising past attacks, but also other content explicitly justifying future attacks.[354] The UK appellate court did not specifically analyze the posts praising past attacks when confirming that over 800 links were "terrorist material" and upholding the conviction under ECHR Article 10.[355]

(i) Tweets: Jokes About Long-Ago Attack – Conviction Violates Free Expression

Case: The State v. Cassandra Vera, Spanish Supreme Court, 2018.

Synopsis: In 2018, the Spanish Supreme Court found free expression violated by a conviction for a series of Twitter jokes referencing a decades-old terrorist attack. The Spanish

350 *The State v. Cassandra Vera*.
351 *Hussain v. The Norwegian Prosecution Authority*.
352 *The State v. Cassandra Vera*.
353 *Leroy v. France*; *Tribunal de grande instance de Paris*.
354 "'Father of terrorism' Khuram Iqbal jailed at Woolwich Crown Court," *BBC News*, Sept. 11, 2014. An appellate court confirmed that the materials at issue—including content praising past attacks, such as videos supporting Al-Qaeda attacks on coalition forces in Iraq and Afghanistan—constituted "terrorist material." *R. v. Iqbal*, No. 2014/01692 C5, para 7, 56, (EWCA Crim 2650, 2014).
355 For more detail, see the discussion above under reposting propaganda without sentiment. *R. v. Iqbal*, para 41.

Supreme Court relied on EU Directive 2017/541 for its definition of terrorist provocation,[356] its requirement that penalized conduct create danger of terrorist acts, and its list of contextual factors to be considered (the author, addressee, and context of the message, as well as the significance and credibility of the threat).[357] The court found none of these indicators present, finding no proof that Vera intended, through her jokes about a 40-years-old attack, to justify the original attack or incite new ones.[358]

In 2013, Spanish citizen Cassandra Vera tweeted a series of jokes about a 1973 bombing assassinating Spain's prime minister by Basque armed group Euskadi Ta Askatasuna (ETA).[359] She was convicted of humiliating victims of and glorifying terrorism and sentenced to one year of imprisonment. The Spanish Supreme Court determined that Vera's conviction had violated freedom of expression; her Tweets neither incited violence nor posed a threat to society.[360]

The court limited criminalization of hate speech to when the hate speech incites violence, analyzing the case under the framework of the ECHR's jurisprudence on hate speech imperiling the rights of third parties. The court relied on EU Directive 2017/541 for its definition of terrorist provocation as "glorification and justification of terrorism or the dissemination of messages or images online and offline, including those related to the victims of terrorism as a way to gather support for terrorist causes or to seriously intimidate the population."[361] The court also adopted the directive's requirement that this conduct, in order to be penalized, must create a danger of

[356] The definition: "The offence of public provocation to commit a terrorist offence act comprises, inter alia, the glorification and justification of terrorism or the dissemination of messages or images online and offline, including those related to the victims of terrorism as a way to gather support for terrorist causes or to seriously intimidate the population." Directive (EU) 2017/541, Mar. 15, 2017. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017L0541&from=EN

[357] *The State v. Cassandra Vera*, STC 493/2018 (Spanish Supreme Court, Feb. 26, 2018). Available at: https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2017/04/TS-Penal-26-febrero-2018.pdf.

[358] *Ibid.*

[359] For instance, she tweeted that the ETA paid the minister's ticket to the moon.

[360] The court's holding did not specify what treaty provisions were violated, but Vera's counsel alleged violations of Article 20 of the Spanish Constitution, Article 19 of the Universal Declaration of Human Rights, and Article 11 of the Charter of the Fundamental Rights of the European Union, all of which guarantee freedom of expression. *Cassandra Vera*, 3.

[361] "The offence of public provocation to commit a terrorist offence act comprises, inter alia, the glorification and justification of terrorism or the dissemination of messages or images online and offline, including those related to the victims of terrorism as a way to gather support for terrorist causes or to seriously intimidate the population." Directive (EU) 2017/541, Mar. 15, 2017.

terrorist acts, as well as its list of contextual circumstances to be considered (the author, addressee, and context of the message, as well as the significance and credibility of the threat).[362] The court found none of these indicators present: finding no proof that Vera intended, through her jokes about a 40-years-old attack, to justify the original attack or incite new ones.[363]

Noting that just because speech is unprotected does not mean it may be criminalized, the court found Vera's conviction disproportionate under national and international law. The court's judgement relied on the following factors: the attack in question occurred 40 years prior; the jokes focused on the method of death rather than the victims themselves; many similar jokes had been made by others without resulting in a criminal prosecution; and lastly, the defendant was only 18 years old, which meant by the time she was born the attacks were distant in history.[364]

(ii) YouTube, Myspace: Song Lyrics Valorizing Terrorist Groups – Unprotected Expression

Case: Case of Jose Miguel Arenas, Spanish Supreme Court, 2018.

Synopsis: In this case, the Spanish Supreme Court found a rapper's lyrics, which praised past attacks and explicitly justified future violence, to constitute incitement and justify a sentence of over three years' imprisonment.

In 2012 and 2013, Spanish rapper Jose Miguel Arenas (under the stage name Valtonyc) posted free audio and video versions of songs on his YouTube and Myspace pages.[365] Some lyrics praised a Basque separatist group, Euskadi Ta Askatasuna (ETA) and a Marxist-Leninst group, First of October Anti-Fascist Resistance (GRAPO). These included "ETA is a great nation," "By killing Carrero, ETA did great, fuck speech, all hail explosives" and "It's justified to shoot your boss in the head, or you could always wait for him to be kidnapped by the GRAPO."[366] Other lyrics referenced murders of Spanish politicians and royal family members, such as "Jorge Campos

[362] *Cassandra Vera*, 10.
[363] *Ibid.*, 10.
[364] Ibid., 8–9.
[365] Spanish original: *Case of Jose Miguel Arenas*, No. 79/2018 (Spanish Supreme Court, Feb. 15, 2018). Available at: https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2018/02/Valtronyc-ruling-Supreme-Court.pdf
[366] *Ibid.*, 3-7

deserves a nuclear bomb" and "The King has an appointment in the town square, a rope around his neck."[367] A Spanish court convicted the rapper for "exalting terrorism and humiliating its victims" as well as for "slandering and insulting the Crown," and sentenced him to three and a half years imprisonment.[368]

The Spanish Supreme Court upheld the conviction as a justifiable limitation on the rapper's freedom of expression.[369] The court reasoned that by creating an atmosphere encouraging of terrorist violence, his lyrics constituted incitement to imitate the violent acts of organizations he praised.[370] The court emphasized no actual violence need result from incitement to establish a risk of such violence occurring.

The court cited Article 5 of the 2005 Council of Europe Convention on the Prevention of Terrorism for a definition of public provocation to terrorism which includes indirect advocacy of terrorist acts.[371] The referenced article directs members establish domestic laws criminalizing "public provocation to commit a terrorist offence," defined as "the distribution, or otherwise making available, of a message to the public, with the intent to incite the commission of a terrorist offence, where such conduct, whether or not directly advocating terrorist offences, causes a danger that one or more such offences may be committed."[372]

The Spanish court justified a sentence of imprisonment by reasoning that public incitement to violence threatened others' fundamental rights and so is unprotected by freedom of expression.[373] Notably, when the rapper fled to Belgium, the Belgian Tribunal of Gand (Ghent) denied extradition because under Belgian law, the lyrics did not qualify as incitement to terrorism; the appeal from this decision is still pending.[374]

[367] *Ibid.*, 3-5.
[368] "Case of Jose Miguel Arenas (Valtonyc)," Columbia Global Freedom of Expression Database, 2018.
[369] The court analyzed the case under the framework of its national constitution, Article 19 of the ICCPR, and Article 10 of the ECHR. *Case of Jose Miguel Arenas*, 9, 12.
[370] "Case of Jose Miguel Arenas (Valtonyc);" *Case of Jose Miguel Arenas*, 9.
[371] "Case of Jose Miguel Arenas (Valtonyc)."
[372] Article 5, Council of Europe Convention on the Prevention of Terrorism (2005) CETS 196.
[373] "Case of Jose Miguel Arenas (Valtonyc)." The court referenced ECtHR jurisprudence to support this proposition, such as *Case of Otegi Mondragon and Others v. Spain*, No. 4184/15 (ECtHR, Third Section, Nov. 6, 2018).
[374] "Case of Jose Miguel Arenas (Valtonyc)."

Note: No English translation of the judgement itself is available, limiting our ability to analyze the reasoning in detail.

(iii) Facebook: Praise of Terrorist Attackers – Conviction Would Violate Free Expression

Case: Hussain v. The Norwegian Prosecution Authority, Norwegian Appellate Court, 2015.

Synopsis: In 2015, a Norwegian appellate court affirmed an acquittal of a man charged with inciting terrorist violence by praising terrorist attacks. The court interpreted the criminal terrorism provision narrowly so as not to cover Husain's Facebook posts, in order to avoid violating the international right to freedom of expression. In doing so, the court prioritized content over context: while the posts eulogized the attackers immediately after the attacks and the speaker was a spokesperson for a terrorist group, his posts did not explicitly urge violence. The court declared that incitement should not be inferred from ambiguous text.

In 2013, Norwegian man Arslan Ubaydullah Maroof Hussain made a series of Facebook posts praising terrorist attacks, often within a few days of the attacks themselves. For instance, four days after the Boston Marathon bombings, he posted pictures of the bombers and wrote: "To hell with Boston and may Allah destroy America. . . May Allah SWT reward them!"[375] The day after two men stabbed a British soldier in London, he posted a link to an article about the attack, commenting "Happy news from England" and calling the attackers "our brave brothers."[376] Two days after the Westgate attacks, he posted about one of the attackers: "May Allah SWT give her a high rank in Paradise!"[377] He posted that the killers of five Norwegians in Algeria deserved

[375] Oda Leraan Skjetne, "I dag faller dommen: Ubaydullah Hussain tiltalt for å oppfordre til terrorhandlinger," [Today, the verdict falls: Ubaydullah Hussain charged with soliciting terrorist acts], *Dagbladet*, Oct. 3, 2014. Available at: https://www.dagbladet.no/nyheter/i-dag-faller-dommen-ubaydullah-hussain-tiltalt-for-a-oppfordre-til-terrorhandlinger/60960484

[376] *Ibid.*

[377] *Ibid.*

"Allah's highest reward."[378] Norwegian courts charged him with inciting terrorist violence by praising terrorist attacks, a charge that carries a statutory penalty of up to six years in prison.[379]

A magistrate court acquitted Hussain of the charge.[380]

The Norwegian Bogarting Court of Appeals affirmed, interpreting the criminal terrorism provision not to cover Husain's posts, in order to avoid violating the international right to freedom of expression.[381] Even though Hussain had served as a spokesperson for an Islamist extremist group, and his statements could reasonably be interpreted as eulogies for the terrorist attacks and attackers, the court deemed his statements too ambiguous to constitute incitement. The court explained criminal liability should not be imposed on the basis of abstract inferences "which have not been clearly expressed and therefore cannot with reasonably great certitude by deduced from the context."[382]

Note: No additional detail is available as to the court's reasoning is available because the judgment is only available in Norwegian and access requires paid registration.

(iv) Regional Newspaper: Cartoon on Recent Attack – Justified Interference

Case: Leroy v. France, ECtHR, 2008.

Synopsis: In 2008, the ECtHR held a French fine levied on a cartoonist for publicly condoning terrorism to be consistent with free expression. The court interpreted the content of the cartoon as glorifying and conveying approval of the attacks, and emphasized how the context of when, where, and how the cartoon was published exacerbated its potential to stir up violence.

Two days after September 11, 2001, a Basque weekly newspaper published a French cartoonist and satirist's drawing of the attack on the Twin Towers with the parody advertising

[378] *Ibid.*
[379] Norwegian Penal Code § 147 c, June 21, 2013, Act of 22 May 1902 No. 10. Available at: https://www.unodc.org/res/cld/document/nor/2005/the_general_civil_penal_code_html/The_general_civil_penal_code_2013.pdf
[380] *Hussain v. The Norwegian Prosecution Authority*, Nos. 14-0499903AST-BORG/01, 14-174730AST-BORG/01 (Norwegian Appellate Court, June 22, 2015). Available at: https://lovdata.no/pro/auth/login#document/LBSTR/avgjorelse/lb-2014-49903-2?searchResultContext=1921
[381] "Hussain v. The Norwegian Prosecution Authority," Columbia Global Freedom of Expression Database, 2015.
[382] *Ibid.*

caption, "We all dreamt of it... Hamas did it."[383] French courts convicted the cartoonist for publicly condoning terrorism and fined him 1,500 euros.[384]

The ECtHR held the fine did not violate freedom of expression as enshrined in Article 10 of the ECHR; instead, restriction was justified and pursued the legitimate aim of public order.[385] The court judged that, rather than merely criticizing American imperialism as the cartoonist claimed, the cartoon's content glorified terrorism; the choice of language as positive as "dreamt" conveyed approval of the terrorist attacks and diminished the victims' dignity.[386]

The court also noted the punishment was only a moderate fine when proclaiming it proportionate. However, the court most strongly emphasized that the context of when, where, and how the cartoon was published exacerbated its potential to stir up violence: only two days after the attack, in the politically sensitive Basque Country, and without accompanying language clarifying a benign intent.

Note: No English translation of the judgement itself is available, limiting our ability to analyze the reasoning in detail.

(v) (National Analogue) Facebook: Satirist's Identification with Recent Terrorist – Unprotected by Free Expression

Case: Tribunal de grande instance de Paris (jugement correctionnel du 18 mars 2015), French Trial Court, 2015.

Synopsis: In 2015, a French trial court convicted a satirist of publicly condoning terrorism. While the court's analysis did not explicitly analyze the comedian's freedom of expression, it applied a similar set of factors as used in international law to convict the defendant. The content did not clearly constitute praise for violence, but contextual factors weighed more heavily. These included time, medium, size of audience, and the speaker's history of anti-Semitic remarks. However, due to his role as a satirist, the court suspended his prison sentence.

383 *Leroy v. France*, Legal Summary, No. 36109/03 (ECHR, Feb. 2, 2008). Available at: http://hudoc.echr.coe.int/eng?i=002-1888
384 *Ibid.*
385 ECHR, Article 10.
386 *Leroy v. France*.

On January 9, 2015, French satirist Dieudonné M'bala M'bala posted a subversive reference to two terrorist attacks on his Facebook page. Specifically, he posted: "tonight as far as I'm concerned I feel like Charlie Coulibaly," referencing both the terrorist attacks on Charlie Hebdo, the French satirical magazine, and Amedy Coulibaly, who killed one policewoman and four Jewish people in the days after the Hebdo attacks.[387] The satirist afterwards explained his post expressed his feeling of being "a comedian treated like a terrorist." A French trial court convicted him of publicly apologizing for an act of terrorism via an online communication and gave him a suspended prison sentence of two months.[388]

The court interpreted the comedian's post as self-identifying with a terrorist in a way that trivialized his terrorist acts, and therefore covered by the statute.[389] The content of the post—the provocative amalgamation of the "Je Suis Charlie" symbol of freedom of expression with a terrorist's name—would normally have qualified as satire, a form of artistic expression.

This was outweighed by the context of the message: posted online during public uproar before the victims were even buried. In terms of impact, the court cited offended online reactions to the post and noted the large size of the comedian's regular audience, as well as the mass audience generally available online. Because the speaker had, along with his followers, a history of hostility towards the Jewish community, the court declared he had a heightened responsibility to consider the impact of remarks like his, which trivialized violence against Jewish victims.

However, noting that comedians and satirists, as part of their artistic expression, regularly made provocative, exaggerated remarks, the court merely gave him a "warning penalty" of a suspended two-months prison sentence, instead of the statutory sentence of five years of imprisonment.[390]

d. Praise of ideology, leaders, members

[387] "TGI, Tribunal de grande instance de Paris, jugement correctionnel du 18 mars 2015," Columbia Global Freedom of Expression.
[388] The satirist also had to pay 2,002 euros in damages to the civil parties in the case. *Tribunal de grande instance de Paris (jugement correctionnel du 18 mars 2015)* (High Court of Paris, Mar. 18, 2015). Available at: https://www.legalis.net/jurisprudences/tribunal-de-grande-instance-de-paris-jugement-correctionnel-du-18-mars-2015/
[389] *Ibid.*
[390] The satirist also had to pay 2,002 euros in damages to the civil parties in the case. *Ibid.*

Eight cases surfaced analyzing freedom of expression for criminal defendants who praised or echoed the ideology, leaders, or members of VEOs. Five were penned by the ECtHR. Of the other three cases, all of which found free expression violated, one came from the UN Human Rights Committee, and the last two came from the Turkish Constitutional Court. The majority of the cases (five) arose in Turkey under charges of disseminating terrorist propaganda for the benefit of the PKK or publicly defending an act by the PKK.

Two cases upheld convictions—neither of which involved a sentence of more than two months in prison—as proportionate and justified infringements on free expression under the circumstances. In one of these opinions (from which eight justices dissented) the ECtHR leaned heavily on contextual factors like the speaker's identity as former mayor of a major city currently roiling with unrest.[391] In the other, the ECtHR reasoned in part that no actual conviction had resulted from the long, but ultimately time-barred, prosecution in issue.[392]

Six cases held that associated convictions violated free expression. All six involved prison sentences for statements that criticized national governments for violent crackdowns or shared the organization's goals without endorsing its use of violence.

These courts focused on the actual content of the statements at issue, none of which praised the commission of past crimes, encouraged violence, or endorsed a VEO's illegal methods of threats, coercion, or violence. Instead, the statements either criticized the government's crackdowns on these groups (such as violence in which civilians are caught in the crossfire[393] or use of solitary confinement[394]) or echoed the ultimate political, social, or ideological goals of the groups (such as Korean reunification[395] or a peaceful, negotiated end to armed conflict[396]). One consisted of wearing a red star (a symbol of Communist ideology) to a protest.[397] Praise of leaders which did not relate to their crimes was considered an insufficient ground for criminalizing

391 *Zana v. Turkey*, No. 40984/07 (ECHR Grand Chamber, Nov. 25, 1997).
392 *Fatih Tas v. Turkey* (No. 3), No. 45281/08 (ECHR Grand Chamber, Apr. 24, 2018).
393 *The Case of Ayse Celikel*, App. No. 2017/36722 (Turkish Constitutional Court, May 9, 2019).
394 *Faruk Temel v. Turkey*, No. 45281/08 (ECtHR Second Section, Apr. 24, 2018).
395 *Keun-Tae Kim v. Republic of Korea*, No. 574/1994 (UN Human Rights Committee, Nov. 3, 1998).
396 *The Case of the Academics for Peace*, No. 2018/17635 (Turkish Constitutional Court, Jul. 26, 2019).
397 *Vanjai v. Hungary*, No. 33629/06 (ECtHR, July 8, 2008).

expression,[398] as was praise of leaders that consisted only of using an honorific like "esteemed" to refer to them.[399]

Of the six cases finding violations, four involved defendants who had criticized their national government on matters of public interest; the courts pointed out this made the government's response of criminalization—already disfavored in the case of speech—subject to especially exacting scrutiny.

Four of these eight cases involved significant prison sentences, ranging from one to two years in prison, all of which were found to violate free expression. Only one of the examined court cases approved a sentence longer than two months in prison.[400]

However, lengthy imprisonment for this type of ideological advocacy certainly occurs. For instance, a Kazakh court sentenced a man to six years imprisonment for distributing terrorist propaganda because he tried to convince his friends of the rightness of ISIS's cause via WhatsApp and phone calls, without directly urging them to join or fight for ISIS.[401] The Supreme Court of the UAE found a journalist guilty of attempting to overthrow the state, under a law authorizing imprisonment for anyone publicly declaring animosity to the state or non-allegiance to the regime.[402] UK courts sentenced a man to three and a half years for posting over 800 links to terrorist propaganda on Facebook,[403] including martyrdom videos, which an appellate court

[398] *Lehideux and Isorni v. France*, No. 24662/94 (ECtHR Grand Chamber, Sept. 23, 1998).
[399] *Faruk Temel v. Turkey*, No. 45281/08 (ECtHR Second Section, Apr. 24, 2018).
[400] Spanish original: *Case of Jose Miguel Arenas*, No. 79/2018, (Spanish Supreme Court, Feb. 15, 2018). Available at: https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2018/02/Valtronyc-ruling-Supreme-Court.pdf
[401] Satkangulov's messages included arguing in favor of the need for armed jihad and justifying takfiirism, the murder of so-called apostate Muslims. The court interpreted Satkangulov's messages as disseminating "concepts justifying terrorism and the need for terrorist activity," and therefore found him guilty of distributing terrorist propagenda and indirectly promoting terrorism. The Kazakh court's decision did not mention or analyze any concerns related to freedom of expression. "The State v. Bulat Zhakpbaevich Satkangulov," Columbia Global Freedom of Expression, 2015. *Opinion itself, in Russian*: "в отношении Саткангулова Б.Ж" [*In relation to Satkangulov B.Zh.*], No. 1-406/2015 (Kazakh Court No. 2 of the City of Kostanay, Nov. 18, 2015). Available at: https://causa.kz/2015/11/prigovor-sud-2-goroda-kostanaya-ot-18-noyabrya-2015-goda-1-406-2015-v-otnoshenii-satkangulova-b-zh/
[402] Federal Law No. (7) of 2014, Art. 15; *United Arab Emirates v. Al-Najjar* (UAE Federal Supreme Court, 2014); "UAE v. Al-Najjar," Columbia Global Freedom of Expression, 2014.
[403] "'Father of terrorism' Khuram Iqbal jailed at Woolwich Crown Court," *BBC News*, Sept. 11, 2014.

(upholding the conviction on grounds related to freedom of expression) confirmed constituted "terrorist material."[404]

(i) Leaflets: Echoing Group's Nonviolent Talking Points – Free Expression Violated

Case: Keun-Tae Kim v. Republic of Korea, UN Human Rights Committee, 1998.

Synopsis: In 1998, the UN Human Rights Committee (HRC) held a South Korean conviction of a political party official for distributing material for the benefit of the North Korean regime violated free expression under Article 19 of the ICCPR. The committee found South Korean had authorities failed to establish how the pamphlets, merely by echoing already well-known views of the North Korean regime, posed a risk to national security or benefited that regime.[405] The committee discounted the disent's argument that the pamphlets caused violent protests, determining that Kim's organization of violent protests did not justify separately punishing him for disseminating the pamphlets.[406]

In January 1989, as an official for the National Coalition of Democratic Movement (NCDM), South Korean citizen Keun-Tae Kim produced and disseminated pamphlets advocating for reunification and criticizing South Korea as a military dictatorship under the thumb of foreign allies, which echoed North Korean talking points.[407] At an NCDM meeting, he read the pamphlets aloud and had them distributed to the 4,000 attendees.[408] He was convicted for praising the activities of and producing materials for the benefit of an anti-state organization (the North Korean regime) and ultimately sentenced to two years imprisonment.[409]

For events occurring soon after—between January 1989 and June 1990—Kim was convicted for organizing illegal demonstrations in which participants threw Molotov cocktails, set

[404] However, some of the content the defendant linked to consisted of justifications for future attacks. *R. v. Iqbal*, No. 2014/01692 C5 (EWCA Crim 2650, 2014).
[405] *Ibid.*, 12.4.
[406] *Ibid.*, 12.5.
[407] *Keun-Tae Kim v. Republic of Korea*, No. 574/1994 (UN Human Rights Committee, Nov. 3, 1998). Available at: https://www.refworld.org/cases,HRC,3f588eff7.html
[408] *Ibid.*, 2.1.
[409] *Ibid.*, 2.1–2.2.

cars afire, and injured over a hundred policemen.[410] Kim was later elected to the National Assembly in 1996.[411]

The UN Human Rights Committee (HRC) held Kim's conviction for propaganda violated his freedom of expression under Article 19 of the ICCPR; South Korean authorities failed to establish how the pamphlets, merely by echoing already well-known views of the North Korean regime, posed a risk to national security or benefited that regime.[412]

The committee reasoned that Kim's organization of violent protests did not justify separately punishing him for disseminating the pamphlets.[413] A dissenting member argued that Kim's pamphlet dissemination was what led to the violent protests in question, making the restriction on Kim's free expression legitimate under national security and public order. .[414] Thus, it seems the majority on the committee disagreed with the dissent's claim that the pamphlets caused the protests, and distinguished correlation from causation.

(ii) Published Interview: Supporting Group's Goal – Justified Interference

Case: Zana v. Turkey, ECtHR, 1997.

Synopsis: In an exceedingly close case, the ECtHR held a Turkish conviction for an ambiguous statement of praise for the PKK did not violate freedom of expression by a vote of fourteen to fourteen. This conclusion incorporated the light sentence involved, of only two months imprisonment. It seems likely that had the sentence of imprisonment been significantly longer, such as two years, a majority of judges would have held the punishment disproportionate.

While in prison in 1987, Mehdi Zana, the former mayor of Kurdish city Diyarbakir, told a journalist: "I support the PKK national liberation movement; on the other hand, I am not in favour of massacres. Anyone can make mistakes, and the PKK kill women and children by mistake."[415] The journalist published the interview in a national newspaper. As a result, Turkish courts

410 *Ibid.*, 4.2.
411 *Ibid.*, 9.2.
412 *Ibid.*, 12.4.
413 *Ibid.*, 12.5.
414 *Ibid.*
415 *Zana v. Turkey*, No. 40984/07, ¶ 12 (ECHR Grand Chamber, Nov. 25, 1997).

convicted Zana of publicly defending a criminal act and sentenced him to two months imprisonment.[416]

The ECtHR determined the aims of protecting national security and public safety were legitimately invoked because the applicant's status as former mayor of a major city gave him political standing and the statement was made at a time of unrest, when the majority of the south-east of Turkey was under a state of emergency and the PKK had recently killed people there.[417]

The ECtHR found the interference justified and proportionate by weighing the statement's context more than its content, as well as the lightness of the punishment.[418] The ECtHR admitted the applicant's words, taken in isolation, were ambiguous and contradictory and would not normally amount to an offense. However, given the circumstances of the applicant's status as former mayor of a key city in south-east Turkey at a time when the PKK had been attacking people in south-east Turkey, the ECtHR considered the statement "likely to exacerbate an already explosive situation" in the region.[419]

One dissenting justice found no legitimate aim for interference.[420] Seven dissenting judges considered Zana's sentence disproportionate, given that he disclaimed support for violence in his statement and was an already-imprisoned former mayor rather than a current mayor. They also pointed out Zana's personal history of opposition to violence.[421]

(iii) Book: Depicting Members as Heroes – Prosecution as Justified Interference

Case: Fatih Tas v. Turkey, ECtHR, 2018.

Synopsis: In 2018, the ECtHR held the Turkish government did not violate free expression by charging the publisher and editor of a book of PKK member auto-biographies with disseminating terrorist propaganda. The court found that some of the book's passages, by

416 *Ibid.*, ¶ 26. *See* Turkish Crim. Code, Article 312.
417 *Ibid.*, ¶ 11, 49, 50.
418 *Ibid.*, ¶ 61–62.
419 *Ibid.*, ¶ 59–60.
420 *Ibid.*, Partly Dissenting Opinion of Judge Thor Vilhjalmsson ("The plain meaning of these words is that the applicant has the same opinion as the PKK on the question of the status of the territory where Kurds live in Turkey but he disapproves of the methods used by this organisation.").
421 *Ibid.*, Partly Dissenting Opinion of Judge Van Dijk, Joined by Judges Palm, Loizou, Mifsud Monnici, Jambrek, Kuris, and Levits.

glorifying and praising violence, could reasonably be interpreted as inciting violence. The court interpreted the defendant's role as publisher and editor gave him practical control over the book's contents, making him vicariously liable for the speech of the authors to which he gave an outlet.

In 2003, Turkish national Fatih Taş published a book collecting autio-biographical accounts of PKK members; three passages depicted PKK members as heroes and praised their actions.[422] These passages romanticized deaths of PKK members, spoke of "getting in a good shot," portrayed vehicle destruction as good news, and described armed clashes with security forces as "nice."[423] Turkish authorities charged Taş with disseminating terrorist propaganda.[424] He underwent over seven years of criminal proceedings, ultimately resulting in a quashed conviction and sentence.[425]

The ECtHR held the interference with Fatih Taş' freedom of expression to be legitimate and proportionate. First, some book passages in question were reasonably considered to incite violence, even though Taş did not write those passages.[426] While passages referring to Turkish security forces as "enemies" did not constitute incitement, three passages glorified armed force and praised violence, and so did—or at least were reasonably interpreted by Turkish authorities as incitement.[427]

Second, contextual factors reinforced the court's conclusion. Because Taş owned the publishing company and was editor-in-chief of the collection, he could shape the book's content, and the court subjected him vicariously to the same duty of care as an author. Even though he did not personally associate himself with these views, he gave the speakers an outlet.[428] Third, the interference with Taş's freedom of expression was deemed slight and therefore proportionate because he was not finally convicted or ultimately imprisoned.[429]

422 *Fatih Tas v. Turkey* (No. 3), No. 45281/08 (ECHR Grand Chamber, Apr. 24, 2018). Available at: http://hudoc.echr.coe.int/eng?i=001-182443
423 *Ibid.*, ¶ 16, 34.
424 *See* The Prevention of Terrorism Act (Law no. 3713), § 7(2).
425 *Fatih Tas v. Turkey*, ¶ 18.
426 *Ibid.*, ¶ 29–37.
427 *Ibid.*, ¶ 34.
428 *Ibid.*, ¶ 35.
429 *Ibid.*, ¶ 36.

(iv) Speech: Referring to a Terrorist Leader with an Honorific – Free Expression Violated Case: Faruk Temel v. Turkey, ECtHR, 2018.

Synopsis: In 2018, the ECtHR held Turkish courts violated free expression by convicting a political party official of disseminating terrorist propaganda. The court found that the official's speech did not incite violence but rather criticized the Turkish government, making imprisonment a wholly unnecessary and disproportionate response. The court emphasized that governments must show restraint in criminalizing speech, and that authorities must tolerate criticism of their actions.

In 2003, Turkish political party chairman Faruk Temel read aloud a statement to around 150 people at a party meeting. In addition to denouncing the American intervention in Iraq, Temel criticized the solitary confinement of PKK leader Abdullah Ocalan, whom he referred to using the honorific "sayın," meaning "esteemed."[430] He also condemned solitary confinement in Turkish prisons in general, and asked for general amnesty for prisoners, including Abdullah Ocalon. Turkish courts convicted Temel of disseminating propaganda to the benefit of the PKK by publicly defending the use of violence or other terrorist methods, and sentenced him to a year's imprisonment, later reduced to ten months due to good behavior.[431]

The ECtHR held the Turkish conviction and sentence violated freedom of expression.[432] While acknowledging the pursued aims of protecting national security and public order were legitimate, the court emphasized that as a member of an opposition political party criticizing the government about matters of public interest, Temel was entitled to the strictest supervision of restrictions on his free expression.[433]

The court found that overall, the content of Temel's speech did not encourage violence or constitute hate speech, so there was no pressing social need to interfere with his expression.[434] Furthermore, the court emphasized that governments must show restraint in criminalizing speech,

[430] *Faruk Temel v. Turkey*, No. 45281/08 (ECtHR Second Section, Apr. 24, 2018). Available at: http://hudoc.echr.coe.int/eng?i=001-103141
[431] *Ibid.*, ¶ 34.
[432] *Ibid.*, ¶ 64.
[433] *Ibid.*, ¶ 52, 55, 63.
[434] *Ibid.*, ¶ 62.

and that authorities must tolerate criticism, making imprisonment a wholly disproportionate response.[435]

Note: No additional detail is available as to the court's reasoning is available because the judgment is not available in English.

(v) Newspaper Article Defending Collaboration – Violated Free Expression

Case: Lehideux and Isorni v. France, ECtHR, 1998.

Synopsis: In 1998, the ECtHR held French courts violated free expression by convicting the authors of an advertisement defending a historical national figure who had collaborated with Nazi Germany. The court noted that the content of the advertisement did not praise collaboration but rather omitted mention of the collaboration, as well as the decades that had passed since the collaboration.

In 1984, daily newspaper *Le Monde* published an advertisement defending Marshal Pétain's collaborative policies towards Nazi Germany as an attempt to save France and inviting readers to write to two associations dedicated to trying to overturn Pétain's conviction and death sentence for collaborating with Nazi Germany.[436] In association with the advertisement, two French nationals were convicted of defending war crimes or collaboration with the enemy: authors Jacques Isorni and Marie-François Lehideux, and sentenced to the symbolic payment of one franc to the civil parties in the case.[437]

The ECtHR held the conviction disproportionately interfered with the authors' freedom of expression.[438] The court assessed that the content of the advertisement did not praise any specific pro-Nazi policy, or even collaboration with Nazis in general, so much as praise a man and omit mention of his collaboration with Nazis.[439] Furthermore, the events discussed in the publication

[435] *Ibid.*, ¶ 63.
[436] *Lehideux and Isorni v. France*, No. 24662/94 (ECHR Grand Chamber, Sept. 23, 1998), ¶ 10. Available at: http://hudoc.echr.coe.int/eng?i=001-58245
[437] *Ibid.*, ¶ 43.
[438] *Ibid.*, ¶ 58.
[439] *Ibid.*, ¶ 53.

had occurred over forty years beforehand.[440] The court also noted the seriousness of criminalization of speech when civil remedies were available.[441]

(vi) Petition Criticizing State Actions Against Group – Free Expression Violated

Case: The Case of the Academics for Peace, Turkish Constitutional Court, 2019.

Synopsis: In 2019, the Turkish Constitutional Court held that convictions of academics for signing a petition that criticized Turkish military actions against the PKK violated free expression because they were unnecessary in a democratic society. The court drew bright lines in terms of criminalizable content. It declared a state could not criminalize speech endorsing the ideology or goals of a terrorist organization, and certainly could not criminalize criticism of public authorities; rather, the state could only criminalize speech justifying, praising, or encouraging such an organization's use of terroristic methods. The court also weighed the speakers' roles as academics in their favor, emphasizing the importance of academic freedom for public debate as a means of government accountability.

In 2016, thousands of academics signed a petition entitled "We will not be a party to this crime,'' which criticized Turkish military actions in Kurdish provinces in Southeastern Turkey, called for an end to the massacre of people in the region, and demanded the Turkish government compensate victims and return to the peace process.[442] Ten of the signatories were convicted by Turkish courts for disseminating terrorist propaganda to benefit the PKK; nine received suspended sentences of 15 months imprisonment, while the tenth was imprisoned for two and a half months.[443]

The Turkish Constitutional Court held these convictions violated freedom of expression, as the interference with the academics' freedom of expression was not necessary in a democratic society.[444] The court declared the state could not criminalize speech endorsing the ideology or

[440] *Ibid.*, ¶ 55.
[441] *Ibid.*, ¶ 57.
[442] "The Case of the Academics for Peace," Columbia Global Freedom of Expression, 2019.
[443] *Ibid*.
[444] *See The Case of the Academics for Peace*, No. 2018/17635 (Turkish Constitutional Court, Jul. 26, 2019) [in Turkish]. Available at: https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2019/09/Academics-for-Peace-Constitutional-Court-Judgment-Turkish.pdf

goals of a terrorist organization; rather, the state could only criminalize speech justifying, praising, or encouraging such an organization's use of terroristic methods of threats, coercion, or violence.[445] The court also recalled that criminally punishing criticism of public authorities was particularly suspect, and concluded any punishment of propaganda on the basis of harm to the government's reputation would violate freedom of expression.[446] The court also emphasized the importance of freedom of expression for academics and public debate as a mechanism of public accountability.[447]

Applying these standards, the court emphasized the petition's content could not be considered to encourage terrorist acts or increase the risk of people committing them, so could not fall within the meaning of the law criminalizing terrorist propaganda.[448] The court dismissed the government's argument that the academics' failure to call on the PKK to end the violence might justify the convictions.[449] The government's case for incitement rested solely on its assertion that a PKK leader had called for Turkish intellectuals to act, but it failed to even present any specific evidence that the academics had been following the PKK's orders when making their statement.[450]

Note: No English translation of this decision is available, precluding more detailed analysis.

(vii) Dial-in to TV Talkshow: Criticizing State Action Against Group – Expression Violated Case: The Case of Ayse Celikel, Turkish Constitutional Court, 2019.

Synopsis: In 2019, the Turkish Constitutional Court held that a conviction for disseminating PKK propaganda violated free expression under the Turkish Constitution. The court reasoned that since her statement—which expressed distress over Turkish military actions in Kurdish regions—did not incite violence or pose an imminent threat of terrorism, interference was not necessary in a democratic society

445 "The Case of the Academics for Peace."
446 *Ibid.*
447 *Ibid.*
448 *Ibid.*
449 *Ibid.*
450 *Ibid.*

In January 2016, Ayse Celikel, a teacher from the Kurdish region, dialed into a popular television talk show in Turkey to express her distress over the killing of children and civilians and round-the-clock curfews resulting in starvation in her area. Without mentioning the PKK, she urged listeners to "not remain silent," "stay stop," "lend a hand," and not to let "those children die."[451] The Turkish Assize Court convicted her for disseminating propaganda for the PKK and sentenced her to fifteen months of imprisonment.[452]

The Turkish Constitutional Court held Celikel's conviction violated freedom of expression under the Turkish Constitution; since her statement did not incite violence or pose an imminent threat of terrorism, interference with her freedom of expression was not necessary in a democratic society.[453]

The court declared that only statements that encourage or incite violent acts may be criminalized as terrorist propaganda; statements merely reflecting or approving of terrorist organizations' ideology or goals cannot.[454] Celikel spoke on an issue of public interest, which meant the state had a very narrow margin for interference, and criticized the actions of public authorities rather than private individuals, which meant she had a broader license.

In terms of context, the court weighed that Celikel spoke spontaneously, in a time of political crisis, in favor of the state's duty to tolerate her speech.[455] In terms of content, the court emphasized that Celikel's speech did not praise the PKK in general, let alone its members fighting in the armed conflict, and did not incite violence or hatred towards the Turkish security forces.[456] Accordingly, the state failed to meet its burden of establishing a strong likelihood her statement could cause a terrorist act.[457]

Note: No English translation of this decision is available, precluding more detailed analysis.

[451] "The Case of Ayse Celikel," Columbia Global Freedom of Expression Database, 2019.
[452] *Ibid.*
[453] *See The Case of Ayse Celikel*, App. No. 2017/36722 (Turkish Constitutional Court, May 9, 2019) [in Turkish].
[454] "The Case of Ayse Celikel."
[455] "The Case of Ayse Celikel."
[456] "The Case of Ayse Celikel."
[457] "The Case of Ayse Celikel."

(viii) Wearing a Red Star – Conviction Violates Freedom of Expression

Case: Vanjai v. Hungary, ECtHR, 2008.

Synopsis: In 2009, the ECtHR held Hungarian courts violated free expression by convicting a man of using totalitarian symbols in public places for wearing a red star to a peaceful protest. The court emphasized that the red star worn by defendant did not exclusively correspond to totalitarian ideology; prohibitions on symbols must distinguish between potential meanings entitled to the protection of free expression and meanings forfeiting that protection. The court also noted that Hungarian authorities had not pointed to any evidence that publicly wearing a red star had ever caused any public disorder, and that none of the wearer's circumstances suggested he intended to cause any such disorder.

In 2004, Atitila Vajnai, the Vice-President of the left-wing Workers' Party (*Munkáspárt*), attended a lawful demonstration in the center of Budapest at the former site of a statue of Karl Marx and was a speaker at the event.[458] On Vajnai's jacket was a five-pointed red star, which was symbolic of the international workers' movement; a police patrol at the scene asked Vajnai to remove the star, which he promptly did.[459] Nonetheless, criminal proceedings were brought against Vajnai under Article 269/B § 1 of the Hungarian Criminal Code, which outlaws the use of totalitarian symbols in public spaces, and in March 2004 he was convicted.[460] The statutory sentence was a fine. The Court of Justice of the European Union found that the case outside its jurisdiction, and Vajnai's conviction was upheld on appeal to the Budapest Regional Court (*Fővárosi Bíróság*).[461]

The European Court of Human Rights found that Vajnai's conviction for the display of totalitarian symbols violated his freedom of expression under Article 10 of the ECHR, as Hungary's restrictions did not pursue legitimate aims of preventing disorder and protecting the

[458] *Vanjai v. Hungary*, No. 33629/06 (ECtHR, July 8, 2008). Available at: http://hudoc.echr.coe.int/eng?i=001-87404.
[459] *Ibid.,* ¶ 6.
[460] *Ibid.,* ¶ 6-8.
[461] *Ibid.,* ¶ 9-13.

rights of others.[462] Furthermore, they were unnecessary, as Hungarian authorities had not established a "clear, pressing, and specific" social need or requirement for the restriction.[463]

The court emphasized that subject to even the Hungarian government's own admission, a red star does not exclusively correspond to an identification with totalitarian ideology. The court acknowledged the red star symbolized the systematic terror used to implement communism and may create "uneasiness" in people.[464] However, the court articulated the necessity of analyzing context when distinguishing between words which forfeit Article 10 protection and those that do not, and pointed out that the law here made no attempt to distinguish between the protected and unprotected ideas associated with red stars.[465]

The ECtHR also weighed two contextual factors in favor of the wearer's free expression: (1) the speaker's personal history and (2) the lack of any violence resulting from previous alike displays. The court pointed out that Vajnai had worn the star at a lawfully organized, peaceful demonstration in his capacity as the vice-president of a registered, left-wing, political party, with no known intention of defying the rule of law.[466] Furthermore, with regard to preventing disorder in society, the court emphasized the Hungarian authorities had failed to indicate any concrete example where a display of the red star had caused any actual or even remote danger of disorder.

D. National Laws

Treatment of freedom of expression varies dramatically between and within different regions. Accordingly, we analyzed a select set of eight countries and their laws that regulate and criminalize extremist speech and speech supporting terrorist organizations. In particular, Germany, Pakistan, Russia, Australia, Brazil, Kenya, Singapore, and Egypt were selected for analysis, on the basis of their representation of their respective regions, their unique laws in addressing online

[462] *Ibid.*, ¶ 58.
[463] *Ibid.*, ¶ 51.
[464] *Ibid.*, ¶ 57.
[465] *Ibid.,* ¶ 53-54.
[466] *Ibid.,* ¶ 53.

extremist or terrorist speech, and/or the enforcement of their national laws in practice. We found the following high-level trends:

1) Several of the surveyed countries, including Germany, Russia, Australia, Brazil, Kenya, and Singapore, rely on legislation designed for regulating "offline" speech as the primary vehicle for regulating online speech; in contrast, Pakistan and Egypt have implemented extremism laws specifically designed to address online activity.
2) While all of the surveyed countries have specialized counter-terrorism laws, passed in the aftermath of the September 11 attacks to respond to jihadist terrorism, these countries use a mixture of laws to address both domestic extremist groups as well as jihadist terrorist organizations. However, amidst the growth of right-wing extremism in their state and/or region, Australia and Germany have passed legislation specifically targeting right-wing speech online.
3) Of the countries surveyed, Pakistan, Russia, Brazil, and Kenya have been criticized for overly vague statutory definitions of terrorism, while those four countries as well as Australia and Singapore have similarly vague definitions of expression that constitutes incitement to acts of terrorism. This has led to the restriction of speech that under international law would not be considered incitement to extremism or terrorism.
4) Starting with the passage of Germany's NetzDG in 2018, several countries, including Pakistan, Russia, Australia, Kenya, Singapore, and Egypt, have adopted legislation that puts more responsibility on content providers to proactively limit potentially offending speech and consequently could encourage the over-restriction of expression. While many of these laws penalize violations by civil sanctions, some countries like Australia and Egypt hold content providers criminally liable for extremist content that is not removed.

The following country-by-country analysis of national laws helped determine the trends identified above.

1. Germany

Germany is a salient example of a recent response to the growth and proliferation of speech on the Internet and social media as well as a changing response to the evolving nature of online extremism. While Germany has a history of both left-wing[467] and right-wing extremism,[468] Germany's main anti-terrorism and extremism laws were implemented in response to 9/11 attacks and growth of jihadist terrorism.[469] As right-wing terrorism has increasingly proliferated in Germany, however, the national government has applied both existing German law and new regulations to restrict extremism online.[470]

For instance, Section 3 of Germany's Vereinsgesetz (Association Law) has been used against both Islamist and far right extremists, outlawing both ISIS in 2014[471] and far-right extremist groups like the neo-Nazi group Combat 18 in Jan 2020 both offline and online;[472] it has also been used in the past against left-wing extremism that proliferated in Western Europe during the Cold War.[473] In contrast, in response to growing threats from right-wing extremists online, the German government recently approved the *Entwurf eines Gesetzes zur Bekämpfung des Rechtsextremismus und der Hasskriminalität*, or Draft Act on Combating Right-Wing Extremism and Hate Crime (hereinafter Draft Act), under which social media companies must proactively

467 "RAF-Serie (II): Der Showdown: Dann Gibt Es Tote,'" *Der Spiegel*, Sept. 16, 2007. Available at: https://www.spiegel.de/spiegel/print/d-52985281.html.
468 Peter Maxwill. "Der Anschlag Auf Einen Schwarzen in Hessen Muss Uns Aufrütteln," *Der Spiegel*, July 24, 2019. Available at: https://www.spiegel.de/panorama/justiz/waechtersbach-in-hessen-schuesse-auf-eritreer-wir-muessen-reagieren-a-1278790.html.
469 "EU and Member States' Policies and Laws on Persons Suspected of Terrorism-Related Crimes." European Parliament Committee on Civil Liberties, Justice, and Home Affairs, Dec., 2017.
470 Keiran Hardy, "Countering Right-Wing Extremism: Lessons from Germany and Norway," *Journal of Policing, Intelligence and Counter Terrorism* 14, no. 3 (Feb. 2019): 262–79. Available at: https://doi.org/10.1080/18335330.2019.1662076.
471 "De Maizière Verbietet Betätigung Der Terrororganisation 'Islamischer Staat' in Deutschland," *Bundesministerium des Innern, für Bau und Heimat*, Sept. 26, 2017. Available at: https://www.bmi.bund.de/SharedDocs/pressemitteilungen/DE/2014/09/verbot-islamischer-staat.html.
472 Jasper von Altenbockum, and Justus Bender, "Seehofer Bans Right-Wing Extremist Group 'Combat 18 Germany,'" *Archyde*, Jan. 23, 2020. Available at: https://www.archyde.com/seehofer-bans-right-wing-extremist-group-combat-18-germany.
473 Russel Miller, "Balancing Liberty and Security in Germany," *Journal of National Security Law & Policy* 4 (2010).

report illegal content.[474] Despite criticisms that the Draft Act incentivizes the over-reporting and likely restriction of expression,[475] its approval by the government accentuates its willingness to directly address growing online extremism with new legislation in addition to its adaptation of existing legislation to new online developments.

Perhaps the most notable of Germany's laws is the Network Enforcement Act (Netzwerkdurchsetzungsgesetz, hereinafter NetzDG), which was passed in June 2017 in response to the handling of criminal content on social networks.[476] NetzDG forces online platforms with more than two million or more registered German users to remove "obviously illegal" posts within twenty-four hours of notification or risk fines up to €50 million. In order to determine whether an act is "illegal," the NetzDG refers to the Criminal Code, in particular to the provisions on dissemination of propaganda material or use of symbols of unconstitutional organizations, encouragement of the commission of a serious violent offense endangering the state, commission of treasonous forgery, public incitement to crime, incitement to hatred, and defamation, among others.

NetzDG and the closely related Draft Act only impose intermediary liability, civilly sanctioning the content providers that serve the offending content to the web page. Violation of these laws brings civil charges, though involving substantial financial penalties. However, since the passage and implementation of NetzDG, similar legislation has proliferated across the globe.

While NetzDG is focused on civil intermediary liability of technology companies, countries have introduced similar bills that not only tighten civil intermediary liability but also introduce individual criminal responsibility for hosting supposedly extremist content.[477] For

[474] *Entwurf eines Gesetzes zur Bekämpfung des Rechtsextremismus und derHasskriminalität*, March 2020. Available at: https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/RefE_BekaempfungHatespeech.pdf?__blob=publicationFile&v=1.

[475] "German Draft Law on Combating Right-Wing Extremism and Hate Crimes Raises Serious Free Expression and Privacy Concerns," Center for Democracy and Technology, March 18, 2020. Available at: https://cdt.org/insights/german-draft-law-on-combating-right-wing-extremism-and-hate-crime-raises-serious-free-expression-and-privacy-concerns/.

[476] "Germany Implements New Internet Hate Speech Crackdown," *DW.com*, Jan. 2018. Available at: https://www.dw.com/en/germany-implements-new-internet-hate-speech-crackdown/a-41991590.

[477] Jacob Mchangama and Joelle Fiss, "The Digital Berlin Wall: How Germany (Accidentally) Created a Prototype for Global Online Censorship," *Justitia*, Nov. 2019. Available at: http://justitia-int.org/wp-

instance, the Philippines' Anti-Fake Content Act[478] authorizes the Department of Justice to order the removal of information "reasonably" believed to be "false or… tend to mislead the public," establishing a criminal penalty of imprisonment in addition to civil penalties of between 300,000 to 2,000,000 Philippine pesos for the publisher of the offending content.[479] Similarly, Australia's amendment to the Australian criminal code "hold[s] content service providers criminally liable for failure to remove 'abhorrently violent' material expeditiously;"[480] this will be further explored below in the section on Australia.

Altogether, Germany serves as an example of a country with a strong legal history of balancing human rights and mitigation of extremist threats in the latter half of the twentieth century, and as such has successfully applied "offline" laws, such as the Vereinsgesetz, to instances of online speech and association. Nonetheless, it exemplifies the trends of introducing new legislation to address the growth of far-right extremism as well as imposing tighter regulations on social media companies that potentially could restrict freedom of expression.

2. Pakistan

Pakistan serves as an example of a state with relatively recent laws that specifically regulating extremist expression online. Pakistan has historically relied on the Pakistan Telecommunication Act of 1996,[481] which provides the authority to regulate telecommunications as well as the rights to set rules on "enforcing national security measures,"[482] and the Prevention

content/uploads/2019/11/Analyse_The-Digital-Berlin-Wall-How-Germany-Accidentally-Created-a-Prototype-for-Global-Online-Censorship.pdf.

[478] Senate Bill No. 9, filed on July 1, 2019 (*an Act Prohibiting the Publication and Proliferation of False Content on the Philippine Internet, Providing Measures to Counteract its Effects and Prescribing Penalties Therefore*). Available at: https://senate.gov.ph/lis/bill_res.aspx?congress=18&q=SBN-9

[479] Joseph Johnson, "The Filipino Anti-False Content Bill: Fake News and Free Expression," OHRH, Oct. 27, 2019. Available at: https://ohrh.law.ox.ac.uk/the-filipino-anti-false-content-bill-fake-news-and-free-expression/.

[480] Criminal Code Amendment (Sharing of Abhorrent Violent Material) Bill, 2019. Available at: https://parlinfo.aph.gov.au/parlInfo/download/legislation/bills/s1201_first-senate/toc_pdf/1908121.pdf;fileType=application%2Fpdf.

[481] Pakistan Telecommunication (Re-Organization) Act, 1996 (with 2006 amendments). Available at: https://www.pta.gov.pk/assets/media/telecom-act-170510.pdf.

[482] *Ibid.,* Art. 57 (ag).

of Electronic Crimes Ordinance,[483] which, despite being created to prevent the use of the Internet for terrorist propaganda and harassment, was seen as imposing significant restrictions on human rights. After the Crimes Ordinances lapsed in 2009, Pakistan had no laws to prevent and prosecute specifically cybercrimes.[484] However, in the aftermath of a 2014 Taliban attack on an elementary school in Peshawar,[485] the Pakistani government drafted and passed the Prevention of Electronic Crimes Act of 2015 (hereinafter PECA).[486]

PECA, which went into effect in 2016, has been roundly criticized for its vague language defining terrorism, as well as the extent to which supposedly non-terrorist crimes, like causing sectarian hate, qualify as cyber-terrorism.[487] In particular, PECA includes in the definition of terrorism actions that advance "inter-faith, sectarian or ethnic hate"[488] as well as granting the Pakistan Telecommunications Authority (hereinafter PTA) unilateral power in determining unlawful online content under Section 37.[489] These vague and broad standards were demonstrated in *Awami Workers Party v. Pakistan Telecommunication Authority,* wherein the PTA blocked the website of the Awami Workers Party under Section 37 of PECA.[490] While the Court ruled against the PTA's application of PECA and corresponding violation of the Pakistani Constitution's protection of due process, the very initiation of such a case emphasizes the broad power of PECA in limiting expression, even when not seen as traditional extremist speech.[491]

[483] Prevention of Electronic Crimes Ordinance 2007, Available at: http://www.pakistanlaw.com/electronic_prevention_ord.pdf
[484] F. Mohammed, "PECA 2015: A Critical Analysis of Pakistan's Proposed Cybercrime Bill." *Journal of Islamic and Near Eastern Law*, 15 (2016). Available at: https://escholarship.org/uc/item/14x2s9nr.
[485] Declan Walsh, "Taliban Besiege Pakistan School, Leaving 145 Dead," *The New York Times*, Dec.16, 2014. Available at: www.nytimes.com/2014/12/17/world/asia/taliban-attack-pakistani-school.html.
[486] Prevention of Electronic Crimes Act, 2016. Available at: http://www.na.gov.pk/uploads/documents/1470910659_707.pdf.
[487] Eesha Arshad Khan. "The Prevention of Electronic Crimes Act 2016: An Analysis," *Lums Law Journal*, Shaikh Ahmad Hassan School of Law, Nov. 1, 2019.
[488] Prevention of Electronic Crimes Act, 2016, § *48*.
[489] *Ibid.*
[490] *Awami Workers Party v. Pakistan Telecommunication Authority*, Writ Petition No. 634 of 2019 (Islamabad High Court, Sept. 12, 2019). Available at: https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2019/10/Order-in-AWP-Website-Blocking-Case.pdf.
[491] *Ibid.,* ¶ 2-3.

Pakistan has also enacted the Anti-Terrorism Act of 1997 (hereinafter ATA),[492] an act intended to increase the power of law enforcement and efficiently process those accused of terrorism through the judicial system, to regulate online speech. The ATA, already seen of using on overly vague definition of terrorism in prosecuting "offline" cases, has been similarly applied online. This trend is best seen through the case *The State v. Saqlain Haider*, where the Saqlain Haider, a Shia Muslim, was arrested after posting disdainful and hateful messages about the Prophet Mohammed on Facebook.[493] Charged with disseminating material that caused "religious, sectarian, or ethnic hatred" (Section 11-W) as well as "committing acts likely to cause hatred" (Section 9),[494] Haider was sentenced to thirteen years in jail in a trial widely criticized by human rights organizations.[495] Indeed, *The State v. Saqlain Haider* set precedent in ensuring the enforcement of the ATA online despite being designed for offline applications; furthermore, the case continues the ATA's trend of increasingly broad enforcement.[496]

Even more recently, Pakistan's government released the Citizens Protection (Against Online Harm) Rules on January 21, 2020,[497] which ostensibly aims to prevent extremist and terrorist speech online as well as, more generally, "hate speech and harassment."[498] However, the act establishes stricter civil regulations on social media companies further than that of NetzDG, requiring the decryption of data for government supervision,[499] the proactive blocking of content related to "terrorism, extremism, . . . and national security,"[500] and the removal within twenty-four

[492] Anti-Terrorism Act, 1997. Available at: https://www.ilo.org/dyn/natlex/docs/ELECTRONIC/81777/88943/F435058093/PAK81777.pdf.
[493] *The State v. Saqlain Haidler* (Pakistani Anti-Terrorism Court, Nov. 21, 2015); "The State v. Saqlain Haider." Global Freedom of Expression, Oct. 16, 2014.
[494] *Ibid.*
[495] Imran Gabol, "Pakistani Shia Man Jailed for 13 Years for Facebook 'Hate Speech,'" *Dawn*, Nov. 24, 2015. Available at: https://www.dawn.com/news/1221725.
[496] "Terror on Death Row: The Abuse and Overuse of Pakistan's Anti-Terrorism Legislation," Justice Project Pakistan and Reprieve, Dec. 2014. Available at: http://www.jpp.org.pk/wp-content/uploads/2017/01/2014_12_15_PUB-WEP-Terror-on-Death-Row.pdf.
[497] Citizens Protection (Against Online Harm) Rules, 2020. Available at: https://moitt.gov.pk/SiteImage/Misc/files/CP%20(Against%20Online%20Harm)%20Rules%2C%202020.pdf.
[498] Michael Karanicolas, "Newly Published Citizens Protection (Against Online Harm) Rules Are a Disaster for Freedom of Expression in Pakistan," Information Society Project, Yale Law School, Feb. 29, 2020. Available at: https://law.yale.edu/newly-published-citizens-protection-against-online-harm-rules-are-disaster-freedom-expression
[499] Citizens Protection (Against Online Harm) Rules, §6.
[500] *Ibid.,* § 4(4).

hours of content deemed illegal by either the National Coordinator or Pakistan Telecommunications Authority.[501] The enforcement of such rules coincides with a growing trend to require the proactive regulation of extremist speech.

Ultimately, Pakistan serves as an example of a country broadening the bounds of content determined as "extremist" and enacting legislation similar — or in this case, more extreme than – – Germany's NetzDG. Notably, in contrast to many of the countries examined in this report, Pakistan's legal basis for addressing extremist speech online comes from laws passed within the last four years that regulate platforms' actions.

3. Russia

Russia has generally relied on laws designed for the regulation of "offline" extremism for addressing online extremist speech. Historically, cases involving online extremist or terrorist speech were addressed through Article 282 of the Russian Criminal Code,[502] which prohibits the "incitement of hatred" or maligning of individual(s) based on their sex, race, religion, etc. as well as the threat of violence or incitement by an organized group,[503] or the 2002 Federal Law No. 114-FZ "On the Counteraction of Extremist Activity,"[504] which prohibits terrorist and extremist activity and any material support to them. Rarely, Russian Courts will cite Article 354.1 of the Russian Criminal Code,[505] which punishes denying the Nuremberg trial's findings and spreading false information about the USSR in World War II.

The Russian Criminal Code and Federal Law No. 114-FZ have seen a wide range of interpretation and broad application by Russian Courts. For instance, in the 2015 cases *The Case of Chesnokov A.V.*[506] and *The Case of Yevgeniy Kort*,[507] the Court found the dissemination of Nazi

[501] *Ibid.,* § 4.
[502] Russian Criminal Code, Article 282. Available at: https://www.wto.org/english/thewto_e/acc_e/rus_e/WTACCRUS48_LEG_6.pdf
[503] *Ibid.*
[504] Federal Law on Counteracting Extremist Activity. Available at: http://host.uniroma3.it/progetti/cedir/cedir/Lex-doc/Ru_Ext-2002.pdf
[505] Russian Criminal Code, § 354.1.
[506] *The Case of Chesnokov A.V.*, I-33/2015 (Russian First Instance Court, July 8, 2015); "The Case of Chesnokov A.V." Global Freedom of Expression, June 16, 2015.
[507] *The Case of Yevgeniy Kort*, No.01-0354/2016 (Russian First Instance Court, Nov. 3, 2016); "The Case of Yevgeniy Kort." Global Freedom of Expression, June 16, 2015.

imagery online justified sanction under Article 282 for association with neo-Nazism and incitement of hatred, and the two men were punished accordingly. These cases are typical applications of hate speech laws directed at preventing incitement by extremist organizations.

However, recent cases have seen the application of laws intended to prevent extremist speech to instances of expression that are protected under international law. In 2016, *The Case of Vladimir Luzgin* dealt with the application of Article 354.1 of the Russian Criminal Code.[508] Luzgin shared an article that discussed a Ukranian nationalist and his alleged cooperation with Nazi Germany, which the Court found as false under expert testimony.[509] Consequently, Luzgin was fined RUB 200,000 in a decision widely decried by human rights advocates as punishing historical debate and free expression as incitement to violence associated with Nazism.[510] Similarly, in the aftermath of Russia's annexation of Crimea, the Consumer Rights Protection Society, a Russian NGO, was sanctioned under Section 280(1) of the Russian Criminal Code, which outlaws public speech that could undermine the territorial sovereignty of Russia, for its extremist and inciting action of referring to Crimea as "occupied territory."[511] The application of articles dealing with extremism in the Russian Criminal Code to prevent unfavored speech emphasizes both Russia's use of traditional "offline" legislation to prosecute online crimes as well as its application to cases that generally fall beyond the purview of terrorism-related or extremism-related cases.

Similarly, Federal Law No. 114-FZ, enacted in response to the growth of jihadist terrorism,[512] has been repurposed in recent years to sanction speech that would seem to fall outside

[508] Russian Criminal Code, § 354.1.
[509] *The Case of Vladimir Luzgin,* No. 2-17-16 (Russian First Instance Court, June 30, 2016). Available at: https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2017/04/luzgin-prigovor.pdf.
[510] Tanya Lokshina, "Online and On All Fronts: Russia's Assault on Freedom of Expression," *Human Rights Watch*, Aug. 3, 2017. Available at: https://www.hrw.org/report/2017/07/18/online-and-all-fronts/russias-assault-freedom-expression.
[511] *The Prosecutor General of Moscow v. The Russian Consumer Rights Protection Society*, Case No. 2-5911/2015 ~ M-6472/2015 (Russia First Instance Court, Aug. 19, 2015. Available at: https://zamoskvoretsky--msk.sudrf.ru/modules.php?name=sud_delo&srv_num=1&name_op=case&case_id=264341814&result=1&delo_id=1540005&new=
[512] Mariya Omelicheva, "Russia's Counterterrorism Policy: Variations of an Imperial Theme," *The Palgrave Handbook of Global Counterterrorism Policy*, 2009, 515–33. Available at: https://doi.org/10.1057/978-1-137-55769-8_23.

the bounds of terrorism or extremism. For instance, in *The Case of Ayupov R.N.*, a video produced by Ayupov that criticized local law enforcement was banned as an extremist content.[513] When Ayupov challenged its censorship in court, the Court found that the prohibition of the video was legal under Art. 13 of Federal Law No. 114-FZ on the basis of being "extremist speech" as determined by expert findings and the lack of justifying evidence in the video.[514] Likewise, in *LLC SIBFM v. Roskomnadzor*, the online news site SIBFM was sanctioned for publishing an article that included a "digitally manipulated image of Jesus Christ" in addition to Vladimir Putin and Alexander Puskin.[515] The Court's ruling accentuated that the display of such imagery "b[ore]the signs of extremism," which justified its censorship under Federal Law No. 114-FZ.[516] Ultimately, these cases, which go against international consensus on freedom of expression, emphasize the broadening reach of both the Russian Criminal Code and Federal Law No. 114-FZ in addressing online crime and targeting speech not traditionally seen as extremist.

Recently, on March 18, 2019, the Russian government amended Article 15-3 of the Federal Law on Information, Information Technology and Protection of Information to advocate for the removal of "unreliable information"[517] by the watchdog agency the *Roskomnadzor*. Under the new law, online publications, which includes social media websites, must delete information identified by the *Roskomnadzor* or face being blocked from the Russian Internet, a requirement that has been compared to a stricter version of Germany's NetzDG.[518] The government also enacted an additional amendment that allows for blocking access to content that "displays obvious disrespect for society, the state . . . or the bodies exercising state power in the Russian Federation" as well as anything that might "[offend] human dignity and public morality."[519] Indeed, the implementation

513 "The Case of Ayupov R.N," Global Freedom of Expression, June 16, 2015.
514 *LLC SIBFM v. Roskomnadzor*, No 4ra/5-5022/2016, (Judicial Collegiate for Administrative Cases of Moscow City Courts, June 6, 2016). Available at: http://media-pravo.info/case-resolution/view/id/2026.
515 "LLC SIBFM v. Roskomnadzor." Global Freedom of Expression, Feb. 22, 2013.
516 *Ibid.*
517 Federal Law of 18.03.2019 No. 31-FZ On Amendments to Article 15-3 of the Federal Law On Information, Information Technologies and on Information Protection.
518 Oreste Pollicino, "Fundamental Rights as Bycatch – Russia's Anti-Fake News Legislation," Verfassungsblog, March 28, 2019. Available at: https://verfassungsblog.de/fundamental-rights-as-bycatch-russias-anti-fake-news-legislation/.
519 Federal Law of 18.03.2019 No. 31-FZ.

of these laws has been criticized for their vagueness in comparison to laws like NetzDG; the Russian laws define misinformation to include "obvious disrespect" to authorities, punish site operators broadly rather than limiting access to the specific unlawful content, and impose significantly more extreme fines.[520]

Altogether, Russia's national laws addressing online extremist speech are generally grounded in its existing criminal code and define extremist speech expansively. Russia's recent laws resembling NetzDG also reflect the trend of requiring proactive content monitoring from social media entities, which may have a chilling effect on expression more generally online.

4. Australia

While Australia's laws that combat extremist speech were written prior to the growth of the modern Internet and large social media companies, they have nonetheless found ready application in online speech; even provincial laws have been applied in such a manner. A prescient example is the case *Blair Cottrell v. Erin Ross,* wherein three individuals staged a "mock beheading" to protest against the building of a mosque and posted it on the far-right nationalist Facebook page, United Patriots Front.[521] Despite the three men's argument that their speech was protected under the Australian Constitution, the court found their video lacked "a legitimate purpose" in its aims and was therefore in violation of the Racial and Religious Tolerance Act, a law from Victoria province that criminalizes incitement on the basis of race.[522] This case, which marked the first in Australia to deal directly with hate speech, emphasized even the local restrictions on online speech in Australia.

More broadly, Australia's restrictions on extremist speech, both online and offline, derive largely from the Australian Criminal Code of 1995,[523] which criminalizes membership in and activities involving extremist organizations largely in Sections 101 and 102.[524] The most clarifying

[520] Pollicino, "Fundamental Rights as Bycatch – Russia's Anti-Fake News Legislation."
[521] *Cottrell v. Ross*, AP-17-2306 (County Court of Victoria, 2019). Available at: https://www.countycourt.vic.gov.au/files/documents/2019-12/cottrell-v-ross-2019-vcc-2142.pdf
[522] *Ibid*, ¶ 125.
[523] Australian Criminal Code Act, 1995. Available at: https://www.unodc.org/res/cld/document/criminal-code-act-1995_html/Criminal_Code_1995_Vol2.pdf
[524] *Ibid.,* § 101-102.

examples of the application of the Criminal Code come from the cases *R v. Vinayagamoorthy and Ors* and *R v. Lodhi.* In *R v. Lodhi*, an Australian man in his mid-30's, was arrested in October 2003 after collecting maps of the Australian electrical supply system and making efforts to obtain explosives in preparation for a terrorist act. Under Sections 101.4-101.6 of the Criminal Code, Lodhi's actions constituted possessing materials for and in preparation of an act of terror and therefore sufficiently warranted sanction. Likewise, in *R v. Vinayagamoorthy and Ors*, three Tamil Australians raised funds for the Liberation Tigers of Tamil Eelam (LTTE), under the guise of fundraising for tsunami relief and were charged with membership in a terrorist organization, providing funds to a terrorist organization, and providing support to a terrorist organization, in violation of Sections 102.3, 102.6(1), and 102.7(1) of the Criminal Code, respectively.[525] These two cases together demonstrate how the Code has been used by courts in both preventative as well as responsive actions against extremist activity in Australia.[526]

The method by which the Criminal Code addresses extremist activity has likewise evolved over time. In 2005, the Criminal Code was amended by Australia's Anti-Terrorism Act, which redefined terrorist acts in Section 100.1 to a significantly broader definition despite criticism from human rights groups.[527] More recently, in response to the terrorist attack in Christchurch against two mosques,[528] the Australian government amended the Criminal Code, which criminalized the sharing of "abhorrent violent behavior" online, as well as held Internet companies liable for failing to remove abhorrent content quickly. Most notably, content providers can face criminal charges as well as the traditional civil penalties.[529] This legislation, which builds upon Australia's legal

[525] *R v. Vinayagamoorthy and Ors* [2010] VSC 148 (Supreme Court of Victoria, Mar. 31, 2010).
[526] Tamara Tulich, "Prevention and Pre-Emption in Australia's Domestic Anti-Terrorism Legislation," *International Journal for Crime, Justice and Social Democracy* 1, no. 1 (May 2012). Available at: https://doi.org/10.5204/ijcjsd.v1i1.68.
[527] *A Human Rights Guide to Australia's Counter-Terrorism Laws*, Australian Human Rights Commission, 2008. Available at: https://www.refworld.org/pdfid/4a2e2d3e2.pdf
[528] Jacob Mchangama and Joelle Fiss, "The Digital Berlin Wall: How Germany (Accidentally) Created a Prototype for Global Online Censorship," *Justitia*, Nov. 2019. Available at: http://justitia-int.org/wp-content/uploads/2019/11/Analyse_The-Digital-Berlin-Wall-How-Germany-Accidentally-Created-a-Prototype-for-Global-Online-Censorship.pdf.
[529] Criminal Code Amendment (Sharing of Abhorrent Violent Material) Bill, 2019. Available at: https://parlinfo.aph.gov.au/parlInfo/download/legislation/bills/s1201_first-senate/toc_pdf/1908121.pdf;fileType=application%2Fpd

history of holding content providers liable for material on their platforms,[530] has been compared to NetzDG and likewise criticized for potentially limiting legitimate speech and incentivizing over-censorship,[531] but is notably harsher for its ability to impose criminal sanctions.

Altogether, Australia's laws limiting extremist expression online reinforce the trend of using existing "offline" legislation to combat online extremism, as well as the implementation of NetzDG-like content provider liability in response to the growth of far-right extremism.

5. Brazil

In South America, Brazil has been proactive in implementing new forms of restrictions on extremist speech. Historically, Brazil's government has dealt with extremism involving racial animus more than jihadist terrorism.[532] In this regard, Brazil has relied on the Brazilian Constitution as well as a mixture of statutes that sanction discrimination against individuals.[533]

These Brazilian laws, typically used for "offline" restriction of extremist or hate speech have been applied online, often reflecting the country's broad protections for "human dignity." For instance, in *Lancellotti v. Facebook*, the Court ruled that under the Brazilian Constitution, freedom of speech must be balanced against an individual or group's "right to dignity."[534] Indeed, similar patterns of legal application have suggested broad protection for "human dignity" over expression, including in cases regarding the distribution of *Mein Kampf*[535] and biblical protest signs.[536] A case that more directly implicates extremist speech is a case involving Antônio Donato Daudson Peret (alias Donato di Mauro) and Marcus Vinicius, known neo-Nazis who filled their social media

530 *Voller v. Nationwide News Pty Ltd, Fairfax Media Publications Pty Ltd, and Australian News Channel Pty Ltd.,* NSWSC 766 (Supreme Court of New South Wales, June 24, 2019). Available at: https://www.caselaw.nsw.gov.au/decision/5d0c5f4be4b08c5b85d8a60d.
531 Mchangama. "The Digital Berlin Wall."
532 Ronaldo Porto Macedo Junior, "Freedom of Expression: What Lessons Should We Learn from US Experience?" *Revista Direito GV* 13, no. 1 (2017): 274–302. Available at: https://doi.org/10.1590/2317-6172201711.
533 *Ibid.*
534 *Lancellotti v. Facebook*, Case No. 0003142-11.2013.8.19.0209 (Court of Appeals of the State of Rio de Janeiro, Nov 30, 2016). Available at: https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2017/06/Ac%C3%B3rd%C3%A3o.pdf.
535 Case No. 0030603-92.2016.8.19.000 (Criminal Court of Rio de Janiero, Feb. 3, 2016).
536 Judgement No. 0045315-08.2011.8.26.0506 (São Paulo Court of Justice, Dec. 10, 2015).

profiles with Nazi and racist messages, photos and videos.[537] They were accordingly sanctioned under a racial/religious tolerance Law 7.716 / 1989 Paragraph 20, which provides for imprisonment for those who "practice, induce or incite discrimination or prejudice based on race, color, ethnicity, religion or national origin." Therefore, In the months preceding the Rio Olympics, Brazil enacted a new anti-terrorism law that redefined terrorism and established punishments for the promotion of terrorism and terrorist financing.[538] Brazilian authorities first applied the law in July 2016 to dismantle a loose, online, pro-ISIS network called "Defenders of Shariah," which accessed websites linked to ISIS and made "pro forma" declarations of allegiance to the group via social media, prior to the Olympics in an operation named "Operation Hashtag."[539] Several UN Special Rapporteurs criticized the law's overly broad definition of terrorism. The law defines terrorism as an "attack against a person, through violence or serious threat, motivated by political extremism, religious intolerance, ethnic, racial and gender prejudice or xenophobia, with the objective of provoking generalized panic."[540] The rapporteurs warned that this definition might be "used to target civil society, silence civil society, ... and criminalize peaceful activities in defence of minority, religious, labour and political rights."[541] Recently, Brazil's Parliament has been considering PL 9555/2018, which would criminalize praise or incitement of terrorism, impose penalties for up to eight years in prison, and potentially increase punishment for offending content widely spread across the Internet.[542]

[537] Ilana Belfer et. al, "Brazilian Skinheads Sentenced to 8 Years for Nazi Propaganda," *Jewish Telegraphic Agency*, May 11, 2016. Available at: https://www.jta.org/2016/05/11/global/brazilian-skinheads-sentenced-to-8-years-for-nazi-propaganda.
[538] Law No. 13,260, March 16, 2016. Available at: http://www.planalto.gov.br/CCIVIL_03/_Ato2015-2018/2016/Lei/L13260.htm.
[539] "Brazil arrests 10 for 'amateur' terror plot against Olympics," *Reuters*, July 2016. Available at: https://www.reuters.com/article/us-olympics-rio-security-operations/brazil-arrests-10-for-amateur-terror-plot-against-olympics-idUSKCN10121E
[540] "Brazil: Counterterrorism Bill Endangers Basic Rights," *Human Rights Watch*, June 8, 2020. Available at: https://www.hrw.org/news/2015/11/13/brazil-counterterrorism-bill-endangers-basic-rights.
[541] "Brazil Anti-Terrorism Law Too Broad, UN Experts Warn," OHCHR, Nov. 4, 2015. Available at: https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=16709&LangID=E.
[542] "Brazil in 2019: Free Speech and Privacy in the Crosshairs. What Are the Threats?" Electronic Freedom Frontier, Jan. 2019. Available at: https://www.eff.org/deeplinks/2019/01/brazil-2019-free-speech-and-privacy-crosshairs-what-are-threats.

Ultimately, Brazil provides an example of a country that applies both existing and relatively recent law to extremist speech online. It similarly follows the trend of broadly applying laws aimed at preventing extremist speech, though it has a long history of doing so even in "offline" cases.[543]

6. Kenya

Kenya's anti-extremism laws are grounded in its history with terrorism, which include the Norfolk Hotel bombing,[544] the U.S. Embassy bombing,[545] and the Paradise Hotel bombing.[546] After these tragedies, the Kenyan government established legal mechanisms for addressing the new extremist threat, passing the Suppression of Terrorism Bill 2003 and the Prevention of Terrorism Bill in 2012.[547] Upon the initial passage of the 2003 Terrorism Bill, human rights groups questioned the necessity of the legislation, noting that it did not meaningfully change the manner by which terrorism was punished but instead significantly expanded the definition of terrorism,[548] to the point of including even "bar brawls . . . or street violence."[549] Similarly, the 2012 Terrorism Bill was criticized for phrases like "reasonable grounds to believe,"[550] which could provide opportunities for the government to detain individuals on the basis of little evidence.[551]

The role of online speech in Kenyan law has evolved dramatically over time. Kenya's primary law for regulating online speech is the Information and Communication Act of 1998,

543 Ronaldo Porto Macedo Junior, "Freedom of Expression: What Lessons Should We Learn."
544 Raymond Muhula, "Kenya and the Global War on Terrorism: Searching for a New Role in a New War," *Africa and the War on Terrorism*, ed. John Davis (Burlington, VT: Ashgate Publishing Company, 2007).
545 Johnnie Carson, "Kenya: The Struggle Against Terrorism," *Battling Terrorism in the Horn of Africa*, ed. Robert I. Rotberg (Washington, D.C.: Brookings Institution Press, 2005).
546 *Ibid.*
547 Charles Lenjo Mwazighe, *Legal Responses to Terrorism: Case Study of the Republic of Kenya*, Naval Post Graduate School, 2012.
548 *Ibid.*
549 "East African Law Society Statement on Kenya's Draft Anti Terrorism Bill," May 29, 2003. Available at: http://www.legalbrief.co.za/article.php/story=20030529104589999
550 Billow Kerrow, "Anti-Terrorism Bill Should be Aligned with the Constitution," Standard Digital, Sept. 23, 2012. Available at: http://www.standardmedia.co.ke/?articleID=2000066749&story_title=Columns:%20Anti-terrorism%20Bill%20should%20be%20aligned%20with%20Constitution
551 Mwazighe, *Legal Responses to Terrorism*.

wherein Section 29 criminalizes sending content or a message that is "grossly offensive."[552] Similarly, Section 35(3c) of the 2012 Terrorism Bill allows for the limitation of "freedom of expression . . . and opinion to the extent of preventing the commission of an offence under the Act," even for expression that occurs online.[553] However, despite the expansion of regulation of online expression and the vague definition of terrorism in the 2012 Terrorism Bill, Kenya's moderation of online content remains very much in flux. In general, Kenya's judiciary system has ruled on a series of cases that have overturned government regulations on expression or expanded expression deemed permissible under the law. For instance, in the case *Andama v. Director of Public Prosecutions*,[554] the Nairobi High Court ruled that restricting the publication of "obscene information in electronic form" — in this case, several vulgar tweets — was a violation of freedom of expression.[555] Interestingly, the court emphasized that the offending legislation came from a time when "government was used to forms of communication that it could easily access and control."[556] Similarly, the cases *Alai v. Attorney General*[557] and *Andare v. Attorney General*[558] have restricted the government's ability to use laws like the Information and Communication Act to widely suppress dissent and criminalize offensive statements, respectively.

In June 2017, Kenya's government passed the "Guidelines for Prevention of Dissemination of Undesirable Bulk Political SMS and Social Media Content via Electronic Communications Networks,"[559] which required social media companies to "pull down accounts used in

[552] Kenya Information and Communications Act. Available at: https://www.unodc.org/res/cld/document/ken/1930/information-and-communications-act_html/Kenya_Information_and_Communications_Act_2_of_1998.pdf.
[553] "The Internet Legislative and Policy Environment in Kenya," Kenya Human Rights Commission, Jan. 2014. Available at: https://www.sbs.ox.ac.uk/cybersecurity-capacity/system/files/The%20ICT%20Legislative%20and%20Policy%20Ennvironment%20in%20Kenya_0.pdf.
[554] *Andama v. Director of Public Prosecutions*, Petition No. 214 of 2018 (Nairobi High Court, July 31, 2019). Available at: https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2019/11/Andama-v-DPP.pdf.
[555] "Andama v. Director of Public Prosecutions." Global Freedom of Expression.
[556] *Andama v. Director of Public Prosecutions* ¶ 52.
[557] *Alai v. Attorney General,* Petition No. 147 of 2016 (High Court of Nairobi, Apr. 26, 2017). Available at: https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2017/05/Alai-v-Attorney-General-Judgment.pdf.
[558] *Andare v. Attorney General*.
[559] *Guidelines for Prevention of Dissemination of Undesirable Bulk Political SMS and Social Media Content via Electronic Communications Network*s, National Cohesion and Integration Commission of Kenya, July 2017. Available at: https://ca.go.ke/wp-content/uploads/2018/02/Guidelines-on-Prevention-of-Dissemination-of-

disseminating undesirable political contents on their platforms" or face civil penalties.[560] These guidelines, hypothesized to have been directly inspired by Germany's NetzDG,[561] specifically address political speech rather than targeting violations like misinformation, hate speech, or extremist content. Consequently, these new limitations have had a chilling effect on freedom of expression more generally.[562]

All in all, an examination of Kenya's laws reveals changing approaches in addressing online speech, with overly broad legislative definitions of terrorism held in check by the Kenyan judicial system. Nonetheless, Kenya similarly reflects the more global trend of holding content providers and other online intermediaries responsible for preventing state-condemned expression on their platforms.

7. Singapore

Singaporean antiterrorism law is distinct from other surveyed countries in its interventionist nature; that is, Singapore's government generally preemptively makes arrests and detentions to mitigate threats far in advance. This is legally supported by the Internal Security Act (hereinafter ISA),[563] which allows for the detention and arrest of suspects "without warrant or judicial review."[564] While the law has generally been used for clear security threats like "espionage, subversion and terrorism,"[565] it has also been applied more generally to activities perceived as threatening the Singaporean government or rule of law.[566]

Undesirable-Bulk-and-Premium-Rate-Political-Messages-and-Political-Social-Media-Content-Via-Electronic-Networks-1.pdf.

560 Mchangama, "The Digital Berlin Wall."

561 *Ibid.*

562 "Kenya," Freedom House, last updated May 18, 2020. Available at: https://freedomhouse.org/country/kenya/freedom-net/2019

563 Internal Security Act, current as of Aug. 26, 2020. Available at: https://sso.agc.gov.sg/Act/ISA1960

564 Senia Febrica, "Securitizing Terrorism in Southeast Asia: Accounting for the Varying Responses of Singapore and Indonesia," *Asian Survey* 50, no. 3 (2010): 569–90. Available at: https://doi.org/10.1525/as.2010.50.3.569.

565 Damien Cheong, "Selling Security: The War on Terrorism and the Internal Security Act of Singapore," *The Copenhagen Journal of Asian Studies* 23, no. 1 (Oct. 2006): 28–56. Available at: https://doi.org/10.22439/cjas.v23i1.691.

566 Garry Rodan, "Preserving the One-Party State in Contemporary Singapore," *Southeast Asia in the 1990s*, Kevin Hewison, Richard Robison and Garry Rodan (eds.), St. Leonards: Allen &Unwin, 1993.

More generally, freedom of expression has been severely limited within Singapore, both in online and online scenarios.[567] These restrictions can be attributed to the strict enforcement of the Sedition Act of Singapore,[568] wherein Section 3 broadly defines sedition, from including promoting "feelings of ill-will… between different races… of the population of Singapore"[569] to more generally raising "discontent… amongst the citizens."[570] The sedition act has been widely enforced for online speech, most notably in the case *Public Prosecutor v. Takagi*, wherein the owner of a Singaporean blog was convicted of violating the Sedition Act for provoking "hatred against foreigners in Singapore."[571] More recently, the Sedition Act and ISA have been used several times to address the growth of jihadist terrorism inside of Singapore, sanctioning individuals posting support for ISIS and other armed insurgencies online, including a seventeen-year-old youth.[572]

Singapore has likewise followed the trend of passing legislation placing legal responsibility on content providers to proactively prevent and restrict access to supposedly violating content. In May 2019, Singapore passed the Protection from Online Falsehoods and Manipulation Bill (hereinafter Protection Bill),[573] which "permits a single government minister to… order the 'correction' or removal of such content where… doing so is in the public interest."[574] Most recently the Protection Bill, which saw its first application in November 2019 over criticism of a state-

[567] Kenneth Roth, "World Report 2020: Rights Trends in Singapore," Human Rights Watch, Jan.14, 2020. Available at: https://www.hrw.org/world-report/2020/country-chapters/singapore.
[568] Sedition Act, current as of Aug. 26, 2020. Available at https://sso.agc.gov.sg/Act/SA1948
[569] *Ibid.,* § 3(1e)
[570] *Ibid.,* § 3(1d)
[571] *Public Prosecutor v. Takagi*, MAC 903124-2015, MAC 903125-2015, MAC 903126-2015, MAC 903127-2015, MAC 903128-2015, MAC 903129-2015, MAC 903130-2015, MAC 903131-2015 (Singaporean District Court of the Criminal Justice Division, Mar. 23, 2016). Available at: https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2018/07/PP-v-Ai-Takagi-Judgment.pdf.
[572] "Update on Actions Taken Under the Internal Security Act," MHA, July 10, 2019. Available at: https://www.mha.gov.sg/newsroom/press-release/news/update-on-actions-taken-under-the-internal-security-act.
[573] Protection from Online Falsehoods and Manipulation, Bill No. 10/2019. Available at: https://sso.agc.gov.sg/Bills-Supp/10-2019/Published/20190401?DocDate=20190401&ViewType=Advance&Phrase=Protection+from+Online+Falsehoods+and+Manipulation&WiAl=1
[574] Kenneth Roth, "World Report 2020: Rights Trends in Singapore," Human Rights Watch, Jan. 14, 2020. Available at: https://www.hrw.org/world-report/2020/country-chapters/singapore

owned investment fund,[575] has focused on addressing the spread of misinformation as a result of the COVID-19 pandemic.[576] The broad application of the Protection Bill, which allows for significantly more discretionary power relative to NetzDG,[577] demonstrates the Singaporean government's emphasis on intermediary liability and wide discretionary ability to limit speech.

Altogether, Singapore's approach to mitigating online extremist speech is relatively unique, as it relies on extremely discretionary existing laws regulating 'offline" speech for mitigating online speech in a likewise unreserved manner. Similarly, it follows the trend more generally of increased pressure on online content providers in regulating content that appears on their platform.

8. Egypt

Egypt has a history of conflict rooted in religious extremism, starting with the banning of Islamist movements in 1954 after the attempted assassination of President Nasser by the Muslim Brotherhood and extending through cycles of violence to the present day.[578] As a result of the 1958 State of Emergency Act, Egypt has often been in a "state of emergency" that has enabled prolonged government restriction of fundamental freedoms like expression and assembly,[579] including a 31-year period between 1981 and 2012[580] as well as since 2017.[581]

[575] Adam Taylor, "First Target of Singapore's 'Fake News' Law Is Facebook Post That Alleged Failed State Investment in Restaurant," *The Washington Post*, Nov. 27, 2019. Available at: https://www.washingtonpost.com/world/2019/11/25/first-target-singapores-fake-news-law-is-facebook-post-that-alleged-failed-state-investment-salt-bae/

[576] Shibani Mahtani, "Singapore Introduced Tough Laws against Fake News. Coronavirus Has Put Them to the Test," *The Washington Post,* March 16, 2020. Available at: https://www.washingtonpost.com/world/asia_pacific/exploiting-fake-news-laws-singapore-targets-tech-firms-over-coronavirus-falsehoods/2020/03/16/a49d6aa0-5f8f-11ea-ac50-18701e14e06d_story.html.

[577] Mchangama, "The Digital Berlin Wall."

[578] David Fielding and Anja Shortland, "'An Eye for an Eye, a Tooth for a Tooth:' Political Violence and Counter-insurgency in Egypt," *Journal of Peace Research* 47, no. 4 (2010): 433-47.

[579] "Egypt: A Return to a Permanent State of Emergency?" International Commission of Jurists, June 2018. Available at: https://www.icj.org/wp-content/uploads/2018/09/Egypt-Return-to-State-of-Emergency-Advocacy-Analysis-brief-2018-ENG.pdf

[580] "Egypt State of Emergency Lifted after 31 Years," *BBC News*, June 1, 2012. Available at: https://www.bbc.com/news/world-middle-east-18283635.

[581] "Egypt Amends Emergency Laws amid Coronavirus Outbreak," *Reuters*, Apr. 22, 2020. Available at: https://www.reuters.com/article/us-health-coronavirus-egypt-lawmaking/egypt-amends-emergency-laws-amid-coronavirus-outbreak-idUSKCN22437A.

These offline restrictions have likewise manifested online: the number of Egyptian citizens investigated or punished by the state for online expression has steadily increased since the Arab Spring began in late 2010.[582] One of the main laws that has enabled these restrictions on online expression is the Egyptian Penal Code, which includes several articles that seem to conflict with international law protecting freedom of expression.[583] Examples of such articles include Article 98, which criminalizes the "denigration of religion;" Article 174, which makes it illegal to use language that could incite action to overthrow the government system; and Articles 184 and 185, which sanctions the insult of state institutions and government officials.[584] Most alarmingly, Article 76 of Egypt's Telecommunication Law of 2003[585] has frequently been used to justify restrictions on the vaguely phrased "misuse of social media"; the Telecommunication Law was cited as the underlying justification for the censorship of YouTube in 2013.[586]

More recently, Egypt's government passed the Cybercrime Law of 2018,[587] which was Egypt's first law focused specifically on cybercrime. The Cybercrime Law retains similarities to Germany's NetzDG, including provisions against the spread of misinformation. However, it goes significantly further than NetzDG's, imposing sanctions ranging from civil fines to criminal imprisonment for up to five years for "link[ing] to content that is contrary to public morality."[588] According to the Open Technology Fund, this broad law has contributed to up to 163 violations of digital expression since its passage, with a significant number of cases justified on the basis of the defendant joining a banned group or misusing social media.[589]

582 "Digital Authoritarianism in Egypt: Digital Expression Arrests 2011-2019," Open Technology Fund, Oct. 24, 2019. Available at: https://public.opentech.fund/documents/EgyptReportV06.pdf.
583 Egyptian Penal Code. Available at: https://manshurat.org/node/14677.
584 "Digital Authoritarianism in Egypt: Digital Expression Arrests 2011-2019," Open Technology Fund, Oct. 24 2019. Available at: https://public.opentech.fund/documents/EgyptReportV06.pdf.
585 Telecommunication Regulation Law of 2003, Arab Republic of Egypt. Available at: http://hrlibrary.umn.edu/research/Egypt/Egypt%20Telecommunication%20Regulation%20Law.pdf
586 "Association for Freedom of Thought and Expression (AFTE) v. Mohamed Hamid Salem," Columbia Global Freedom of Expression, Jan. 9, 2019..
587 Anti-Cyber and Information Technology Crimes Law, Law No. 175 of 2018, Arab Republic of Egypt.
588 *Ibid.*
589 "Digital Authoritarianism in Egypt."

Conclusion

These country-by-country analyses demonstrate similarities between several countries in addressing extremist speech online, despite regional and country-level differences. In Brazil and Kenya, for example, interpretations of what is considered extremist speech are mainly generated by the state's judicial system, while the other five countries examined generally rely on statutory definitions of extremism or terrorism. Similarly, there is growing concern that many of these legislative restrictions aid national security — or other governmental priorities — at the expense of freedom of expression more broadly.[590]

Our examination found four primary trends in national laws regulating online expression.

First, countries like Germany, Russia, Australia, Brazil, Kenya, and Singapore, rely on existing legislation designed for regulating "offline" speech for regulating online speech. The exceptions to this are Pakistan and Egypt, whose online cybercrime laws have been implemented and enforced relatively recently.

Second, while all countries have counterterrorism laws adopted with jihadist extremism specifically in mind, in practice these countries use a mixture of laws to address both domestic and/or jihadist extremist groups. However, Australia and Germany have passed legislation specifically targeting right-wing speech online.

Third, Pakistan, Russia, Brazil, and Kenya have been criticized for vague definitions of terrorism in penal codes, and those four countries as well as Australia and Singapore have similarly vague definitions of inciting language under recently passed laws.

Finally, starting with the passage of Germany's NetzDG in 2018, Pakistan, Russia, Australia, Kenya, and Singapore have adopted laws that impose greater liability on online intermediaries, like content providers. More notable, however, is the trend of countries like Australia and Egypt enforcing criminal liability on intermediaries, potentially imprisoning people who host and fail to remove third-party extremist content from online platforms they control.

590 Jack Balkin, "Free speech is a triangle," *Columbia Law Review* 118, no. 7 (2018): 2011-2056.

E. Social Media Content Policies

Social media content policies reveal that most private companies' content moderation standards (although perhaps not their actual practices) go further than many courts and nations have when criminalizing terrorist content. In general, this section explores trends as observed in social media content policies.

Every social media platform we assessed has a policy that prohibits terrorist organizations from using its forums to recruit members or advocate for violence, and extends this prohibition to re-posting terrorist propaganda or sharing content that glorifies terrorist leaders or violent acts or posting terrorist symbols or insignia in a positive light. Some make exceptions for educational or awareness-raising purposes, while placing the burden on the poster to make this context clear. These exceptions reflect a varying acknowledgement across the platforms of the importance of *context* and *proportional infringement* on free expression when users violate their terms of service, though most platform policies reflect that context and intent are essential to consider. Encrypted messaging applications, as well as messaging boards like 4chan, generally have sparser terms of service, with blanket and non-specific provisions that users not use the applications to violate the law.

Together, some of these companies founded the Global Internet Forum to Counter Terrorism (GIFCT) to cooperate on technological solutions like the hash database to help thwart terrorists' use of their services. GIFCT runs the Hash Sharing Consortium most major social media platforms use to collaborate on identifying terrorist content to facilitate automatic removals of that content. Summary statistics on the hash database provided by GIFCT help characterize the universe of terrorist propaganda and the relative prevalence of different types of this propaganda. Of the more than 200,000 unique pieces of content hashed in the database in July 2019, 85.5% were categorized as "Glorification of Terrorist Acts," 9.1% consisted of "Radicalization, Recruitment, Instruction," 4.8% depicted "Graphic Violence Against Defenseless People," and just 0.4% posed an "Imminent Credible Threat."[591]

[591] *GIFCT Transparency Report*, Global Internet Forum to Counter Terrorism (GIFCT), July 2020.

1. Content Policy Summaries

Facebook declares that it does not allow "organizations or individuals that proclaim a violent mission or are engaged in violence to have a presence on Facebook," including terrorist organizations.[592] Facebook removes content that supports or praises groups, leaders or individuals involved in coordinated violence, as well as coordination of support for them.[593] This includes symbols that represent those entities, unless shared in a "context that condemns or neutrally discusses the content."[594] Facebook owns Instagram, which also prohibits support or praise of terrorism, organized crime, or hate groups.[595]

YouTube's community guidelines prohibit praise, promotion, and aid to violent criminal organizations, further disallowing the use of YouTube for recruitment to these organizations.[596] YouTube instructs users not to post content aimed at encouraging acts of violence, praising or justifying organizations' violent acts or depict their symbols in a positive or promoting manner.[597] Specific examples of content not allowed include "[r]aw and unmodified reuploads of content created" by terrorist organizations and "[c]ontent directing users to sites that espouse terrorist ideology, are used to disseminate prohibited content, or are used for recruitment."[598] YouTube's parent company Google states more simply, in its content policies for publishers, that it does not allow "content that incites or glorifies violence," or "content used to threaten or organize violence or support violent organizations."[599]

Twitter does not allow users to "affiliate with [or] promote the illicit activities of a terrorist organization."[600] Specifically, users may not engage on behalf of terrorist organizations, recruit

592 *Community Standards*, Facebook. Available at: https://www.facebook.com/communitystandards/dangerous_individuals_organizations
593 *Ibid.*
594 *Ibid.*
595 *Community Guidelines*, Instagram. Available at: https://help.instagram.com/477434105621119
596 *Violent Criminal Organizations*, YouTube Help. Available at: https://support.google.com/youtube/answer/9229472?hl=en&ref_topic=9282436
597 *Ibid.*
598 *Ibid.*
599 *Content Policies*, Publisher Center Help, Google. Available at: https://support.google.com/news/publisher-center/answer/6204050?hl=en
600 "Terrorism and violent extremism policy," Twitter Help Center. Available at: https://help.twitter.com/en/rules-and-policies/violent-groups

for them, post propaganda to further their stated goals, or use their symbols to promote them. [601] More broadly, Twitter prohibits "content that glorifies acts of violence in a way that may inspire others to replicate those violent acts and cause real offline harm, or events where members of a protected group were the primary targets or victims."[602]

Reddit prohibits content that "encourages, glorifies, incites, or calls for violence or physical harm against an individual or a group of people."[603] It does make some exceptions however, which are explored in the analysis below.

TikTok does "do not allow dangerous individuals or organizations to use our platform to promote terrorism, crime, or other types of behavior that could cause harm."[604] Specifically, TikTok does not allow the posting of "[c]ontent that praises, glorifies, or supports dangerous individuals and/or organizations" or even "[n]ames, symbols, logos, flags, slogans, uniforms, gestures, portraits, or other objects meant to represent dangerous individuals and/or organizations."[605]

Tumblr prohibits content that promotes, encourages, or incites acts of terrorism, including content that "supports or celebrates terrorist organizations, their leaders, or associated violent activities."[606] More broadly, Tumblr instructs users not to "post content that encourages or incites violence, or glorifies acts of violence or the perpetrators." [607]

SnapChat's guidelines contain a blanket statement that prohibits terrorist organizations from using SnapChat and announces zero "tolerance for content that advocates or advances

[601] *Ibid.*
[602] "Glorification of violence policy," Twitter Help Center, as of March 2019. Available at: https://help.twitter.com/en/rules-and-policies/glorification-of-violence
[603] "Do not post violent content," Reddit Account and Community Restrictions. Available at: https://www.reddithelp.com/en/categories/rules-reporting/account-and-community-restrictions/do-not-post-violent-content
[604] Community Guidelines, TikTok. Available at: https://www.tiktok.com/community-guidelines?lang=en
[605] *Ibid.*
[606] Community Guidelines, Tumblr, last modified Jan. 23, 2020. Available at: https://www.tumblr.com/policy/en/community
[607] *Ibid.*

terrorism."[608] More broadly, SnapChat also instructs users not to depict "gratuitous violence," and not to encourage violence.[609]

WhatsApp's terms of service peremptorily require that users not use the service in ways that are "are illegal . . . threatening, intimidating, harassing, hateful, racially, or ethnically offensive, or instigate or encourage conduct that would be illegal, or otherwise inappropriate, including promoting violent crimes."[610]

Telegram's terms of service state only that users agree not to "[p]romote violence on publically viewable Telegram channels."[611] However, following pressure from states,[612] Telegram has announced that its abuse team "actively bans ISIS content" by banning bots and channels.[613] However, Telegram commits that it "will not block anybody who peacefully expresses alternative opinions."[614]

Gab Social "strives to be the home of free speech online" and declares that "[p]olitical speech protected by the First Amendment is welcome."[615] Gab's terms of service do not mention violence or terrorism, but do require users to agree not to violate U.S. law: namely, not to "[u]nlawfully threaten," not to "[i]ncite imminent lawless action," and more generally not to "aid, abet, assist, counsel, procure or solicit the commission of, nor constitute an attempt or part of a conspiracy to commit, any unlawful act."[616]

4chan's rules also do not mention violence or terrorism, but direct that users "not upload, post, discuss, request, or link to anything that violates local or United States law." 4chan directs users not to post troll-like, racist, or grotesque posts outside of "/b/" the "random" image board.

608 Community Guidelines, Snap Inc., copyright 2019. Available at: https://www.snap.com/en-US/community-guidelines
609 *Ibid.*
610 WhatsApp Legal Info, WhatsApp. Available at: https://www.whatsapp.com/legal/
611 Terms of Service, Telegram. Available at: https://telegram.org/tos
612 Amar Toor, "Telegram says it will remove terrorist content after Indonesia threatens a ban," *The Verge*, July 17, 2017. Available at: https://www.theverge.com/2017/7/17/15980948/telegram-indonesia-isis-terrorism-moderation-ban
613 "ISIS Watch," *Telegram*, last edited Dec. 26, 2016. Available at: https://t.me/isiswatch/2
614 Telegram FAQ, Telegram. Available at: https://telegram.org/faq
615 "About," Gab. Available at: https://gab.com/about
616 Terms of Service, Gab, copyright 2019. Available at: https://gab.com/about/tos

2. Policy Exceptions and Analysis

Notably, some platforms provide explicit exceptions to their content rules. Reddit allows some violent content that does not explicitly fall into their defined category (which includes some terrorist social media functions) if it is justified by the user with a reason other than intent to incite or recruit when it is posted. Specifically, they write "we understand there are sometimes reasons to post violent content (e.g., educational, newsworthy, artistic, satire, documentary, etc.) so if you're going to post something violent in nature that does not violate these terms, ensure you provide context to the viewer so the reason for posting is clear."[617] TikTok similarly provides exceptions for certain justifications, writing in their policy that "educational, historical, satirical, artistic, and other content that can be clearly identified as counterspeech or aims to raise awareness of the harm caused by dangerous individuals and/or organizations" are exceptions.[618] It seems that Reddit maintains that their violent content category is always prohibited, but provides exceptions to content that may be in a "grey area" or on the edges of that definition provided that the user includes an explanation, while TikTok allows exceptions for content even if it explicitly violates their definition of violent content.

YouTube's exception takes a middle ground between Reddit and TikTok's policies. YouTube writes in their guidelines that if content related to terrorism or a crime is posted for "an educational, documentary, scientific, or artistic purpose" it may be permissible. They suggest that when a user posts this content they should "be mindful to provide enough information in the video or audio itself so viewers understand the context."[619]

These policies highlight two salient considerations in enforcing content moderation policies for terrorist functions on social media: proportionality and intent. The broader reach of these policies (e.g. prohibiting passive praise and even symbols) reflect a distinction between

617 "Do Not Post Violent Content," Reddit Help, as of Feb. 1, 2020. Available at: https://www.reddithelp.com/en/categories/rules-reporting/account-and-community-restrictions/do-not-post-violent-content.
618 "Tik Tok Community Guidelines," as of Feb. 1, 2020. Available at: https://www.tiktok.com/community-guidelines?lang=en.
619 "YouTube Policies," as of Feb. 1, 2020. Available at: https://www.youtube.com/about/policies/#community-guidelines.

removing content and criminalizing it. A proportional infringement on free expression rights of a user when they post a symbol of a terrorist organization may be removal of the content, but criminalizing it or removing the user from the platform after one offence may be disproportionate. Second, the exception policies included in the guidelines of TikTok, YouTube, and Reddit reflect an understanding that considering intent and context remains essential. It is nearly impossible for platforms to weigh all contextual information and infer intent at the scale and timeline required for consistent content moderation. However, these exceptions illustrate that some platforms broadly agree that intent matters when enforcing policies against terrorist content, especially if the content is intended for educational, artistic, or newsworthy purposes.

VI. SPREADING FALSE INFORMATION ONLINE

This report analyzes ongoing global efforts to regulate and criminalize the spread of false information online. First, it provides an analysis of the extent to which criminalization of spreading false information complies with international law. Second, it gives an overview of how states respond to the spread of misinformation and disinformation, whether through criminalization, civil and intermediary liability, campaigns, or other directives. This paper depicts regional trends in state practice and, using statistical analysis, determines which regions legislate with regard to misinformation and disinformation in an atypical fashion. Third, this paper provides a sampling of how different countries define criminalized sharing of false information, and how national courts have responded to the question of whether or not criminalizing the sharing of false information is compatible with the right to freedom of expression.

The term false information will be used in this paper to cover both misinformation and disinformation. For the purposes of this paper, misinformation is defined as the *unintentional* spread of false information.[620] Thus, misinformation is spread through non-culpable ignorance, negligence, or reckless disregard as to the falsity of the information.

In contrast, disinformation is defined as *intentionally* spreading false information in order to build a certain narrative or achieve a certain overarching purpose.[621] Thus, the initial sharer of disinformation spreads the information knowing that it is false. Disinformation can, for example, layer true and false data, and strategically select or omit information. State-sponsored propaganda is the paradigmatic example of disinformation.

[620] Eliška Pírková, *Fighting misinformation and defending free expression during COVID-19: recommendations for states, Access Now*, 2020, at 2. Also, in Claire Wardle and Hossein Derakhshan, *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making*, Council of Europe, Sept. 27, 2017 (available at https://rm.coe.int/information-disorder-report-Nov.-2017/1680764666), three different types of information are defined. Mis-information is "when false information is shared, but no harm is meant", dis-information is "when false information is knowingly shared to cause harm", and mal-information is "when genuine information is shared to cause harm, often by moving information designed to stay private into the public sphere." The latter category is not used in the present report.

[621] Pírková, at 2-3.

A. Relevant Principles of International Law

There are four key areas where misinformation or disinformation can and do fall under existing legal frameworks for expression: incitement to violence or hate speech, fraud or false advertising, defamation, and 'memory' laws concerning false information about historical occurrences. International human rights law jurisprudence clearly permits the criminalization of false information that rises to the level of fraud, incitement to violence, or hate speech, but prohibits the criminalization of mere defamation. In these four areas, international law has relatively clear fixed standards.

Turning to the criminalization of misinformation or disinformation more broadly, established international legal principles yield less clarity. Entities such as the UN Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, the Organization of American States (OAS) Special Rapporteur on Freedom of Expression, and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information have expressed the clear opinion that general prohibitions on the dissemination of information based on "vague and ambiguous ideas, including 'false news' or 'non-objective information' are incompatible with human rights law and should be abolished."[622]

While the cases on point are relatively few, what international jurisprudence does exist on the topic of generalized false information suggests that the classic three-part test of legal prohibition, necessity, and proportionality still applies to any restriction. International courts applying human rights treaties have not applied this test to hold that the criminalization of misinformation or disinformation *per se* violates the right to freedom of expression contained in their respective treaties. However, they have held that in particular instances the prohibition of

[622] *Joint Declaration on Freedom of Expression and Fake News, Disinformation and Propaganda*, The United Nations (UN), the Organization for Security and Co-operation in Europe (OSCE), the Organization of American States (OAS) the African Commission on Human and Peoples' Rights (ACHPR), 2017.

spreading false information was too vague, or that the speech at issue—being political speech—was too important for the criminal law to stand.

1. Established Frameworks for Specific Types of False Information: Opinions by International Bodies

There are four key areas where misinformation or disinformation can and do fall under existing legal frameworks for expression: incitement to violence or hate speech, fraud or false advertising, defamation, and 'memory' laws concerning false information about historical occurrences. International human rights law jurisprudence permits the criminalization of false information that rises to the level of fraud, incitement to violence, or hate speech, but prohibits the criminalization of mere defamation. In these four areas, international law has relatively clear fixed standards.

First, incitement to violence or hate speech, which often includes false information, may clearly be criminalized under international law. Additionally, knowingly spreading false information with a profit motive generally falls under the framework of fraud, which may also be criminalized under international law.

So-called 'memory' laws which criminalize the spread of false information about past genocides, wars, or historical facts, have a mixed record. International courts have upheld applications of these laws only to the extent that the false statements at issue constitute incitement to violence or hate speech.

Lastly, jurisprudence from international courts clearly holds that mere defamation may not be criminalized. However, state practice shows that defamation frequently is criminalized. This phenomenon is covered in more detail under the chapter on defamation.

a. Incitement to Violence and Hate Speech

Criminalizing false information is permitted under international law insofar it meets the incitement to violence standard of hate speech. For more information, see the chapters on VEOs or Cyberharassment.

b. Fraud or False Advertisement

Criminalization of false information will be in undisputed compliance with international law when it constitutes some form of fraud. For instance, in the context of the COVID-19 pandemic, the UN Special Rapporteur brings attention to "a small industry of hucksters seeking to translate people's desire for a cure into a quick source of profit."[623] He adds that such false claims may be sanctioned as long as the tests of Article 19(3) ICCPR are met.[624] However, as long as there is a profit motive, it seems that the spread of false information—of disinformation, in that case—falls squarely under the classical categories of fraud or false advertisement, which are the realm of criminal law. A different story, for which international law does not have a conclusive answer is when—to follow up with the same scenario—someone falsely claims there is a cure for COVID-19, but they do not do it to profit from that claim. That scenario, not being fraud, does not fit in one of the established frameworks.

c. Defamation

According to international jurisprudence, criminalization of defamation—as opposed to the imposition of mere civil penalties for defamation—violates free expression.[625] However, a substantial number of states have criminalized defamation anyway. This tension is discussed in more detail in the chapter on defamation.

An extra criterion that is relevant when analyzing whether a prohibition of misinformation is compliant with international law is the political nature of the speech. This principle applies in the defamation context,[626] but is also applicable to other categories of speech restrictions, including blasphemy[627] and incitement.[628] If the piece of false information being penalized amounts to political speech, international law accords it heightened protection.

For instance, the Human Rights Committee has advised that "public interest in the subject matter" of a criticism should be recognized as a defense to a defamation charge. Furthermore, the

623 David Kaye, *supra*, at ¶ 46.
624 *Ibid.*
625 *Kimel v. Argentina* (Inter American Court on Human Rights, 2008); *Issa Konate v. The Republic of Burkina Faso*, Application No. 004/2013 (African Court on Human and People's Rights).
626 *See Lingens v Austria*, app. no. 9815/82 (ECtHR 1986).
627 *Wingrove v. UK*, app. no. 17419/90 (ECtHR 1996), at ¶ 58 (holding that an application of a blasphemy law violated free expression).
628 *Faruk Temel v. Turkey*, No. 45281/08 (ECtHR Second Section, Apr. 24, 2018).

committee's guidance states that, with respect to allegedly defamatory comments about public figures, states should avoid penalizing "untrue statements that have been published in error but without malice."[629]

While not every topic debated in public may qualify as political, criticism of government will almost certainly be considered political. For instance, the ECtHR interprets ECHR Article 10 to leave little scope for restricting "political speech or on debate of questions of public interest," while affording a wider margin of appreciation to states regulating free expression on "matters liable to offend intimate personal convictions within the sphere of morals or, especially, religion."[630]

d. Denials of Historical Fact (Memory Laws)

There is another parcel of the false information universe where international law has established guidance: the spread of false information about past genocides, wars, or related historical facts. While concerns about misinformation, disinformation and "fake news" in general have been brought to the forefront of political and legal debate in recent years, so-called "memory" laws have been debated for longer. Thus, international human rights tribunals have been able to set standards to ascertain the compliance of memory laws with human rights law, which mirror standards for hate speech. International jurisprudence from the UN Human Rights Committee and the European Court of Human Rights (ECtHR) considers the ICCPR and ECHR to permit criminalizing false historical claims only if the claims, when taken in context, constitute incitement to hatred or violence. In practice, this means that criminalization of outright Holocaust denials has been upheld as permissible under international law, while false statements of fact on other topics have not.

The UN Human Rights Committee has noted that generally, 'memory laws' tend to violate the international human right to free expression as enshrined in ICCPR Article 19. This committee, which monitors the implementation of the ICCPR by its State parties, issued a comment in 2011 which interprets Article 19 as barring the "general prohibition of expressions of an erroneous

629 Human Rights Committee, *supra*, at ¶ 47.
630 *Wingrove v. UK*, application no. 17419/90 (ECtHR 1996).

opinion or an incorrect interpretation of past events." [631] The comment declares: "laws that penalize the expression of opinions about historical facts are incompatible with […] freedom of opinion and expression."[632]

However, the Human Rights Committee has upheld specific applications of memory laws when the statements criminalized constituted hate speech. In the 1996 *Faurisson* case, the committee held that criminal prosecution of a Holocaust denier was justified insofar as the statements made amounted to anti-Semitic speech.[633] The French law in question criminalized contesting the existence of crimes against humanity, the category of crimes for which Nazi leaders were convicted by the International Military Tribunal at Nuremberg. The defendant publicly denied that Nazi concentration camps used gas chambers to exterminated Jews, implying that Jewish historians had concocted this as a false story. While noting that the abstract French law may certainly be applied in ways that violate the ICCPR, the committee ruled only that this specific application of the law was a permissible restriction of the speaker's free expression because his statements, taken in context, tended to stoke anti-Semitism.

The ECtHR has held some applications of memory laws criminalizing false statements about historical fact—all of which related to Holocaust denials—to be permissible. However, the ECtHR has struck down other applications as violations of free expression; in these instances, the court concluded the criminalized statements did not constitute incitement to hate speech or violence.

In five cases between 1998 and 2015, the ECtHR held cases involving denial of atrocities against Jewish people during the Holocaust inadmissible. In three of these cases, the court held the restrictions permissible under Article 10 because they incited racial hatred and thus were necessary in a democratic society.[634] In two of them, the court held the statements in question ineligible for

631 *General Comment No. 34*, Human Rights Committee, at ¶ 49.
632 *Ibid.*
633 *Faurisson v. France,* Communication No. 550/1993, U.N. Doc. CCPR/C/58/D/550/1993 (UN Human Rights Committee 1996). Available at: http://hrlibrary.umn.edu/undocs/html/VWS55058.htm
634 *See Witzsch v. Germany (no. 1)* (dec.), no. 41448/98 (ECtHR, Apr. 20, 1999); *Schimanek v. Austria* (dec.), no. 32307/96, (ECtHR, Feb. 1, 2000); *Gollnisch v. France* (dec.), no. 48135/08, (ECtHR, June 7, 2011).

protection under Article 10 because they fell afoul of Article 17.[635] ECHR Article 17 declares that actions "aimed at the destruction of any of the rights and freedoms set forth" elsewhere in the ECHR are ineligible for the protections of the ECHR, such as Article 10's protection of free expression. In these cases, defendants' attempts to rehabilitate the Nazi regime by accusing its victims of falsifying history exploited free expression rights for ends contrary to the fundamental values of the ECHR: destroying the rights and freedoms of others.

However, not all purported denials of Nazi crimes automatically justify criminalization at the ECtHR. In *Lehideux and Isorni v. France*, the ECtHR held in 1998 that free expression was violated by the conviction of two defendants for penning a misleading news advertisement under laws criminalizing publicly defense of war crime or crimes of collaboration with the enemy. The defendants had praised a historical figure known for collaborating with the Nazis while omitting the historical facts of collaboration and re-interpreting some of his acts using theories that had been refuted by historians. The court found that the defendants' interpretation of the events in question did "not belong to the category of clearly established historical facts – such as the Holocaust – whose negation or revision would be removed from the protection of Article 10 by Article 17."[636] Accordingly, the ECtHR found that while outright Holocaust denial would justify criminalization, because such anti-Semitism would constitute an attempt to destroy the rights and freedoms of others, advancing unsupported historical theories would not.

In contrast, the ECtHR has assessed applications of memory laws to other historical debates in a highly contextual fashion. The court has repeatedly clarified that its role is not to arbitrate historical truth or falsity,[637] but rather to balance the rights of the speaker and affected identity groups in a case-specific fashion, paying careful attention to "the interplay between the nature and potential effects of such statements and the context in which they were made." [638]

[635] *Garaudy v. France* (dec.), no. 65831/01 (ECHR 2003); *Witzsch v. Germany* (ECtHR, Dec. 13, 2005.
[636] *Lehideux and Isorni v. France*, No. 24662/94, at ¶ 47 (ECHR Grand Chamber, Sept. 23, 1998).
[637] See *Chauvy and Others v. France*, no. 64915/01, § 69 (ECtHR 2004); *Monnat v. Switzerland*, no. 73604/01, § 57, (ECtHR 2006); *Fatullayev v. Azerbaijan*, no. 40984/07, § 87, 22 Apr. 2010; and *Giniewski v. France*, no. 64016/00, § 51 *in fine*, ECHR 2006-I).
[638] *Perinçek v. Switzerland*, no. 27510/08, at ¶ 220 (ECtHR 2015).

For instance, in the 2015 case of *Perinçek v. Switzerland,* the ECtHR held that Switzerland had violated free expression by convicting someone for denying the Armenian genocide, because the defendant's statements did not constitute incitement to hatred or violence when taken in context.[639] The Swiss law in question criminalized "grossly trivializ[ing] or seek[ing] to justify a genocide or other crimes against humanity" on the grounds of race, ethnic origin, or religion. Swiss courts sentenced the defendant to a fine, as well as a suspended sentence for further fines which could be replaced by one month's imprisonment. Despite holding the restriction on the defendant's free expression was prescribed by law and properly sought to achieve the legitimate aim of protecting the rights of Armenians to identity and dignity, the ECtHR concluded Switzerland had failed to establish that the restriction was necessary in a democratic society because the statements did not incite hatred or violence.

The court's analysis of the necessity of criminalizing the defendant's statements was highly contextual, taking into account the following factors: (1) the content and form of his statements (including the words themselves and the position of the speaker), (2) the geographical, historical, and temporal context in which the statements were made, (3) the extent to which the statements had impacted the rights of Armenians, (4) whether or not Switzerland had an international legal obligation to criminalize the statements, (5) the existence or lack of consensus between states party to the ECtHR on the criminalization of denying genocide, and (6) the severity of the restriction.

The ECtHR found a lack of necessity for criminalizing the defendant's denials of the Armenian genocide, applying these factors to reach the following findings. First, statements in question did not call for hatred or intolerance towards Armenians and, entitling them to heightened protection under Article 10, were on a matter of public interest. Second, the context of the statements was "not marked by heightened tensions or special historical overtones in Switzerland," and the statements were made more than 90 years after the historical events in question. Third, the statements did not affect the dignity of Armenians sufficiently to require criminalization, in part because their dissemination was limited (they were made orally at public events rather than printed

[639] *Ibid*, at ¶ 229-241.

and handed out widely) and in part because they blamed external imperialist powers for inciting violence between Turks and Armenians rather than criticizing Armenians. Fourth, parties to the ECHR varied on criminalization of genocide denials, exhibiting a lack of consensus; five did not criminalize any denials of historical events, six only criminalize denials of the Holocaust and other Nazi crimes, and twelve criminalize denials of any genocides. Fifth, no international treaty or customary law principle compelled Switzerland to criminalize genocide denials. Finally, sixth, criminalization is a severe restriction on free expression. As a result of all these factors, the ECHR found the restriction disproportionate and not necessary in a democratic society.

In summary, when it comes to false historical claims, international human rights bodies tend to reach for the tools provided by the hate speech framework. The resulting analysis is highly contextual. If making false statements about historical facts constitute incitement to violence or hatred in the context that they were made, their criminalization will be justified. In contrast, if false statements of historical fact do not incite hatred or violence, their criminalization will violate freedom of expression.

2. Expert Guidance on Criminalization of False Statements in General

In 2017, the UN Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, the Organization of American States (OAS) Special Rapporteur on Freedom of Expression, and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information issued a joint statement on the matter. These officers and bodies do not issue binding opinions, and are not sources of international law, but their interpretation of the law may have persuasive authority. In that joint statement, they expressed the clear opinion that general prohibitions on the dissemination of information based on "vague and ambiguous ideas, including 'false news' or 'non-objective information' are incompatible with human rights law and should be abolished."[640]

[640] *Joint Declaration on Freedom of Expression and Fake News, Disinformation and Propaganda*.

The UN Special Rapporteur also noted that the concept of misinformation is too elusive to be used in law, as it is too susceptible to granting unchecked power to governments to determine what is truth and what is not.[641] Therefore, it seems that the problem lies in that the definition of the prohibited speech, whether it be "misinformation" or "fake news" or any equivalent term, is overbroad and thus may fails the clarity part of the prescribed-by-law requirement for any permissible restriction on free speech. The UN Special Rapporteur also concluded that the way vague prohibitions of false information empower state officials to determine the truth or falsity of such statements also conflicts with the necessity and proportionality tests of Article 19(3).[642]

Moreover, the UN Special Rapporteur went as far as to affirm that the "penalization of disinformation is disproportionate, failing to achieve its goal of tamping down information while instead deterring individuals from sharing what could be valuable information."[643] In a similar vein, after the landmark 2015 defamation opinion of the African Court on Human and Peoples' Rights *Lohé Issa Konaté v. Burkina Faso,* the Special Rapporteur on Freedom of Expression and Access to Information in Africa stated that the decision "will pave the way for the decriminalisation of similar law such as [the] publication of false news."[644]

Therefore, expert opinions on human rights law seem to agree that the prohibition of false information under general definitions is not compliant with freedom of expression. In order for such prohibition to pass muster with international law, it would have to be more concrete, and tied to a specific harm. The question remains if a non-criminal prohibition would be sufficient to mitigate that harm. It if were sufficient, criminalization would not pass the proportionality test under Article 19(3).

641 David Kaye, at ¶ 42.
642 *Id.*, at ¶ 49.
643 *Id.*, at ¶ 42.
644 *See* "African Court Addresses Freedom of Expression in Landmark Judgment," International Resource Justice Center, Feb. 3, 2015. Available at: https://ijrcenter.org/2015/02/03/african-court-addresses-freedom-of-expression-in-burkina-faso-in-landmark-judgment/

3. Decisions by International Courts on the Criminalization of False Information

While the cases on point are few, what international jurisprudence does exist on the topic of generalized false information suggests that the classic three-part test of legal prohibition, necessity, and proportionality still applies to any restriction. International courts applying human rights treaties have not held that the criminalization of misinformation or disinformation *per se* violates the right to freedom of expression contained in their respective treaties. Regional international courts like the ECtHR, East African Court of Justice, and the Court of Justice of the Economic Community of West African States (ECOWAS Court) have held that in particular instances the prohibition of spreading false information was too vague, or that the speech at issue—being political speech—was too important for the criminal conviction to stand.

This section contains the only four international court cases identified which analyze the free expression implications of restrictions on misinformation or disinformation as such. Two from ECOWAS and one from the EACJ dealt with criminal penalties. The fourth, from the ECtHR, dealt with civil suppression. The search yielding these cases consisted of three stages: (1) a systematic search of the ECtHR's HUDOC database with key words such as "misinformation, disinformation, fake news, false information, false news;" (2) a non-systematic search of other international court databases that did not allow keyword or subject searching; and (3) a survey of academic articles and databases, including a comprehensive search of the Columbia Freedom of Expression Database.

Other cases identified via this search were included in preceding or following sections. Most cases found, even those described as misinformation or disinformation cases, ended up being defamation cases or memory law cases instead. These were incorporated in the previous sections on legal frameworks for specific types of false information. Cases by national courts analyzing the free expression implications of false information laws were incorporated in the following section on state practice.

a. Case: Federation of African Journalists (FAJ) and others v. Gambia, ECOWAS Court, 2018.

Synopsis: In 2018, ECOWAS Court found that criminal sanctions imposed by Gambia on a number of journalists violated the first prong of the classic tripartite test for restrictions on free expression, due to the chilling effect of the vague penal code provisions.

In 2018, the Court of Justice of the Economic Community of West African States (ECOWAS Court) found that Gambia's criminalization of false news violated the right to freedom of expression as recognized in the African Charter on Human and People's Rights, the ICCPR, and the Treaty of the Economic Community of West African States. Gambian authorities had arrested the applicants under charges for sedition, defamation, and false news. The ECOWAS Court found that the relevant penal code provisions violated the right to free expression due to vagueness and directed Gambia to amend its laws.[645]

The ECOWAS Court found that the criminal sanctions imposed on the journalists violated the classic tripartite test for restrictions on free expression, due to the chilling effect of the vague penal code provisions. Under the three-part test, restrictions on free expression must be provided for by law as well as necessary in order to meet a legitimate aim: to preserve the rights and reputations of others, or protect national security or public order, or public health or public morals. The court concluded that the laws criminalizing sedition, defamation, and false news were too vague to provide proper notice to citizens, and that the criminal sanctions imposed on the applicants were disproportionate and not necessary in a democratic society.[646]

Specifically, the ECOWAS Court held that the legislative definitions of key terms were too vague to be properly 'provided for by law' under the three-part test. The Gambian provision criminalizing false news provided for the imprisonment of up to two years of anyone who "publishes or reproduces any statement, rumour or report which is likely to cause fear and alarm to the public or to disturb the public peace," while "knowing or having reason to believe" that the statement "is false." The Court found this provision subject to "diverse subjective interpretations"

[645] *Federation of African Journalists (FAJ) and others v. The Gambia* (Community Court of Justice of the Economic Community of West African States Apr. 2018), at 4.
[646] *Id*, at 47.

and therefore too broad, as it failed to give notice to speakers as to when their speech would be legal or illegal.[647]

The court put special emphasis in the chilling effect that the Gambian laws would have on journalism. The court relied on both the UN Human Rights Committee General Comment No. 34 and the U.S. case of *New York Times v. Sullivan*.[648] This U.S. case articulates the way broad or vague prohibitions on speech lead to self-censorship of even legal speech because speakers will seek to steer far clear of the potentially unlawful zone. The court also referenced *Sullivan* when finding that an "erroneous statement is inevitable in free debate, and that it must be protected if the freedoms of expression are to have the 'breathing space' that they need."[649] Ultimately, the ECOWAS court declared that the provisions on sedition, defamation, and false news fail to "guarantee a free press within the spirit" of international treaties.[650]

b. Case: Media Council of Tanzania v. Attorney General, East African Court of Justice, 2019.

Synopsis: In 2019, the East African Court of Justice (EACJ) found that multiple provisions of a Tanzanian criminal misinformation law violated free expression under the African Charter because they failed the first prong of the tri-partite test; they were too vague to be properly prescribed by law. In contrast, the EACJ upheld other criminal misinformation provisions which contained mens rea requirements with respect to the falsity of the information.[651]

In 2019, the East African Court of Justice found that Tanzania's misinformation provisions violated free expression. Three NGOs had brought a challenge to Tanzania's Media Services Act, which the court directed Tanzania to amend in order to comply with free expression under the African Charter on Human and Peoples' Rights.

647 *Id*, at 40.
648 *Id*, at 41.
649 *Id*, at 44.
650 *Id*, at 47.
651 *Media Council of Tanzania v. Attorney General* (East African Court of Justice, March 28, 2019), at ¶ 94. Available at: http//globalfreedomofexpression.columbia.edu/wp-content/uploads/2019/05/Referene-No.2-of-2017.pdf

Tanzania's Media Services Act established the following criminal misinformation offences: for publishing, intentionally or recklessly, false information that threatened the interests of defense, public safety and order, morality, or health of Tanzania (Section 50.1.a); for publishing information "which is maliciously or fraudulently fabricated" (Section 50.1.b); "statements knowingly to be false or without reasonable ground for believing it to be true" (Section 50.1.d); and for sharing false statements likely to cause fear or harm (Section 54).

The East African Court of Justice (EACJ) applied a version of the classic three-pronged test to uphold two sections of the law, including two of the misinformation provisions, while ruling that sixteen others violated free expression.[652] The court articulated the tripartite test as requiring that limitations on free expression must: (1) be prescribed by law in a way that gives clear notice as to what is prohibited, (2) have objectives that are pressing, substantial, and important, and (3) be proportionate to those objectives.

The EACJ found that one of the criminal misinformation provisions in the Media Services Act was too vague and hard for individuals to follow,[653] and thus violated the freedom of expression provisions of the African Charter.[654] This provision, Section 54, criminalized the publication of "any false statement, rumor or report which is likely to cause fear and alarm to the public or to disturb the public peace." The statutory penalty for violations is either four to six years in prison or a fine of over 4,000 USD. The provision contained no *mens rea* requirement, but provided that the speaker's having taken measures prior to publication justifying his reasonable belief that the statement was true would be a defense.

The second misinformation struck down as too vague was Section 50(1)(c), which criminalizes using media services to publish any statement which threatens Tanzanian defense, public safety, or public order, economic interests, public morality or public health, or is injurious "to the reputation, rights and freedom of other persons."

652 The provisions of the Media Services Act held to violate the African Charter were sections 7(3); 19; 20; 21; 35; 36; 37; 38; 39; 40; 50; 52; 53; 54; 58; and 59. The provisions of the Media Services Act held to comply with the African Charter were sections 13 and 14.
653 *Id.*, at ¶ 94.
654 *Id.*, at ¶ 95.

Conversely, the EACJ found that another set of criminal misinformation provisions with *mens rea* requirements as to falsity passed the test and complied with the African Charter, without going into detailed analysis.[655] These provisions, in Section 50(1) of the Media Services Act, all related to publications using media services, with mens rea requirements as to falsity ranging from fraud, malice, intentionality, and recklessness, to—in one instance—"without reasonable grounds for believing [the statement] to be true."[656] Section 50(1) provided for violators to receive sentences of either between three and five years in prison or fines of at least 2,000 USD.

The court did not make any effort to distinguish why the struck provisions were too vague and the upheld provisions were not—instead, the court only said summarily that, when applying the provided-by-law requirement to Section 50, the section seemed "largely unobjectionable," while the struck provisions (Section 54 and Section 50(1)(c) were too vague and broad, failing to enable journalist or other individuals to regulate their conduct.

However, the presence of mens rea requirements as to falsity seemed determinative. The upheld portions all contained *mens rea* requirements as to falsity. In contrast, the one portion of Section 50(1) struck down by the court contained no requirement of falsity or specified level of mens rea: Section 50(1)(c), criminalizing the publication of statements threatening to Tanzania's interests or injurious to the reputation, rights, and freedoms of others). Section 54, which the court struck down, contained a falsity requirement but, despite allowing for reasonable grounds for believing the statement true to serve a defense, did not specify a *mens rea* requirement as to falsity as a prima facia element of the offense.

c. Case: Brzeziński v. Poland, ECtHR, 2019.

Synopsis: In 2019, the ECtHR ruled that Polish courts violated free expression due to the chilling effects of ordering a local candidate for public office to cease distribution of a booklet

655 *Id.*, at ¶ 94.

656 The following subsections upheld by the court all related only to publication using "media services." Section 50(1)(a) criminalized publishing information which is intentionally or recklessly falsified that threatened the interests of defense, public safety and order, economy, morality, or health of Tanzania. Section 50(1)(b) criminalized publishing information "which is maliciously or fraudulently fabricated." Section 50(1)(d) criminalized publishing a statement with knowledge of its falsity or "without reasonable grounds for believing it to be true." Section 50(1)(e) criminalized publishing a statement maliciously or with fraudulent intent "representing the statement as . . . true."

alleged to include false information, publish an apology, and donate to charity. These penalties violated free expression because they were disproportionate and not necessary in a democratic society.[657]

In 2006, Polish local electoral candidate Zenon Brzezinski handed out an election booklet criticizing a mayor for mismanaging water contracts and accusing a councilor of corruption. The mayor and councilor sued Brzezinski, seeking to stop disseminating the pamphlet under Section 72 of the Polish Local Elections Act. This law authorized court orders banning campaign material including false information and required courts to determine whether or not to issue the order no later than 24 hours after a complaint was filed. The trial was held three hours after Brezezinski was notified.

The ECtHR held that while the interference with Brzezinski's free expression was prescribed by law and pursued the legitimate aim of protecting the reputation and rights of others, it was not necessary in a democratic society and disproportionate. The court found that the "cumulative application" of the penalties on Brzezinski had a chilling effect on political speech.

First, the Court reiterated that, for political speech or debate on questions of public interest, Article 10 of the ECHR permitted very little restriction, and that criticism of public officials requires greater leeway than criticism of private individuals. The court found that the speech by Brezinzki, expressing himself as an electoral candidate and opponent of the incumbent mayor, was certainly of public interest.

Furthermore, the Court determined that Polish courts had apparently failed to examine whether or not the brochure's contents had a credible factual basis, refusing to endorse the Polish court's decision to place the burden of proving the booklet's truth on the defendant. The court determined that the language in the booklet was within the limits of "exaggeration and provocation" of normal political debates. As a result, the Polish courts had failed to demonstrate a pressing need for the interference, and the interference was disproportionate.

With respect to the summary nature of the proceedings and time pressure created by the 24-hour statutory deadline for decision, the ECtHR noted that the urgency was justified by the

[657] *Brzeziński v. Poland*, no. 47542/07 (ECtHR, 2019).

need to ensure that "fake news" and other statements which may damage electoral candidate's reputations and may distort the results of an election are rectified as soon as possible. Notably, the *Brzeziński* decision marked the first time the ECtHR judges used the term "fake news," although it was not defined by the court.

Note: No English translation of this decision is available, precluding more detailed analysis.[658]

B. State Practice on Combatting the Spread of False Information

In order to assist in orderly analysis, the 194 countries assessed in this study were organized according to geographic and political region. Considering the location of each nation, the affiliation of each nation to supranational bodies, and international linguistic commonalities, the following ten regions were established: Africa, Asia, the Caribbean, Central and South America, the Commonwealth of Independent States, Europe (non-EU), Europe (EU), the Middle East, North America, and Oceania.

This section begins with a quantitative analysis of these 194 countries, including global and regional metrics of to what extent these countries have legislatively responded to false information, and to what extent these responses have included criminalization. Countries that have criminalized or at least legislatively considered criminalizing misinformation or disinformation constitute, overall, a minority of countries. However, they make up a majority of the majority of countries that have legislatively responded to the spread of false information. These quantitative findings are followed by a qualitative summary of trends observed, and then a sampling of specific laws passed in different regions of the world.

1. Quantitative Analysis

a. Caveats

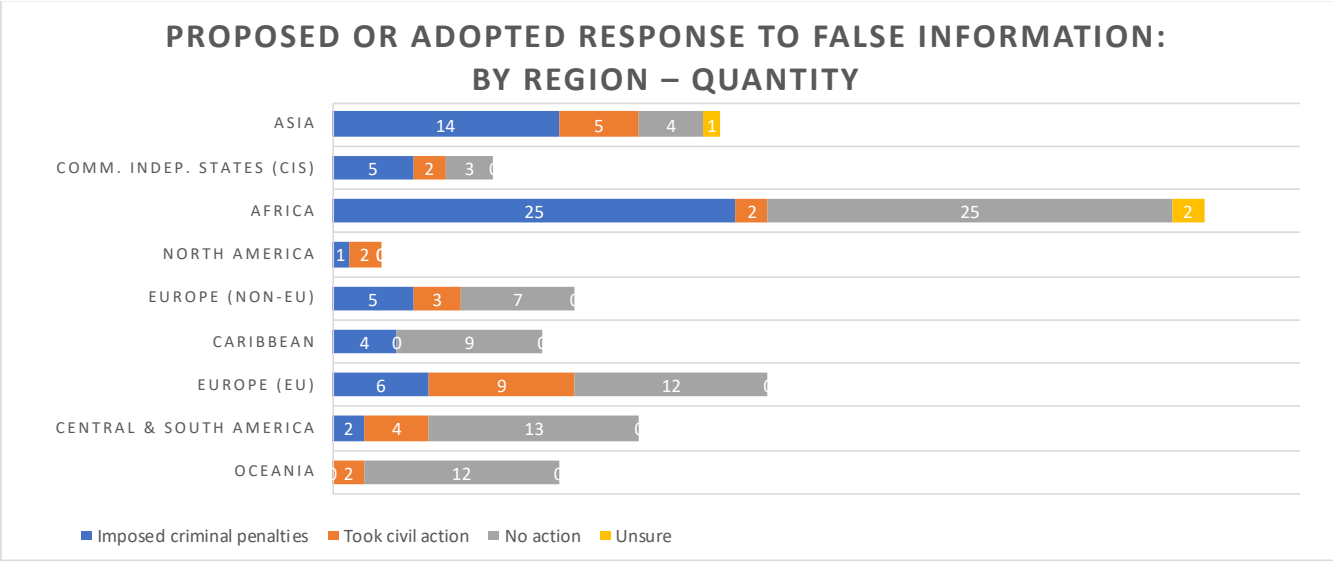
Two key caveats precede this analysis of global and regional trends with respect to the national responses to the spread of false information.

[658] Full text in French available at: http://hudoc.echr.coe.int/eng?i=001-194958

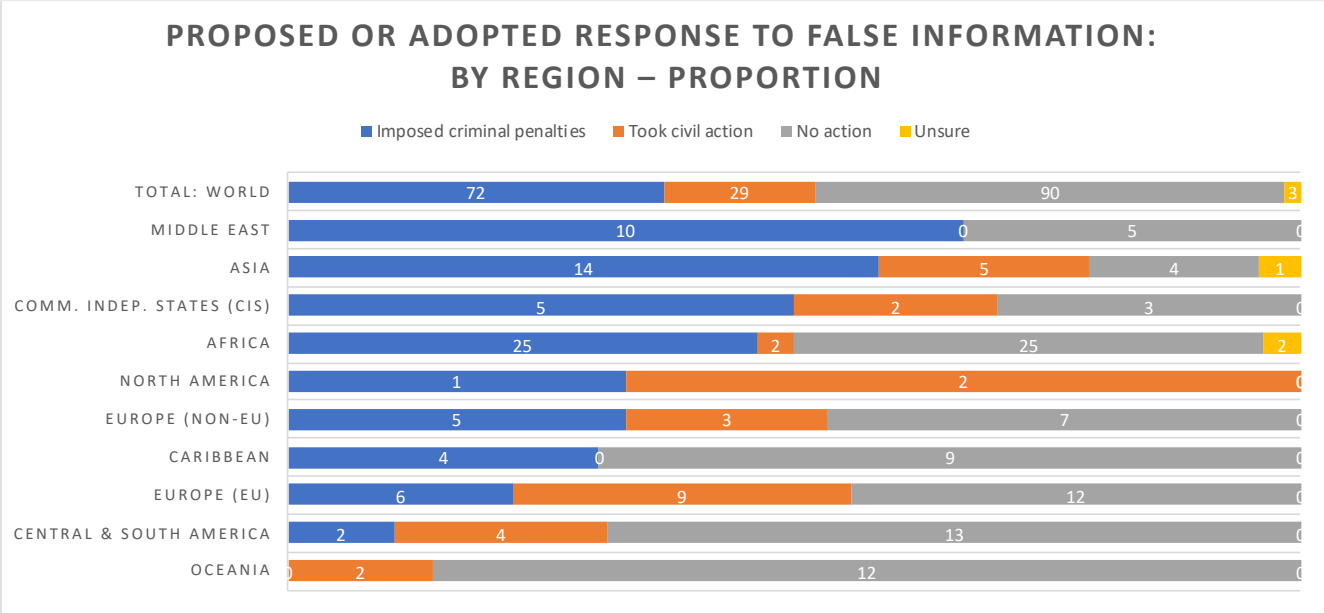
First, most sources consulted were news, legal, and academic materials written in English that were accessible via search engine and database queries in the United States. Access to complete and accurate English information about some nations' laws is limited.

Second, this section does not differentiate between actions that are being processed, passed, implemented, and repealed by governments. The research team only included proposed legislation as an 'action taken or considered' when the team could not confirm whether it had passed or not; if the team could verify that a piece of legislation had not been passed, the team did not include it in the count. As a result, this section prioritizes exhaustivity over currency. Bureaucracies, legislatures, and other governmental entities move at different rates in different countries. Because it is possible that some of the legislation or orders included in this paper's data analysis have been rejected, repealed, amended, or slow-rolled at the time of this paper's writing, the actions included in the regional trend analysis consist of both passed and merely proposed actions.

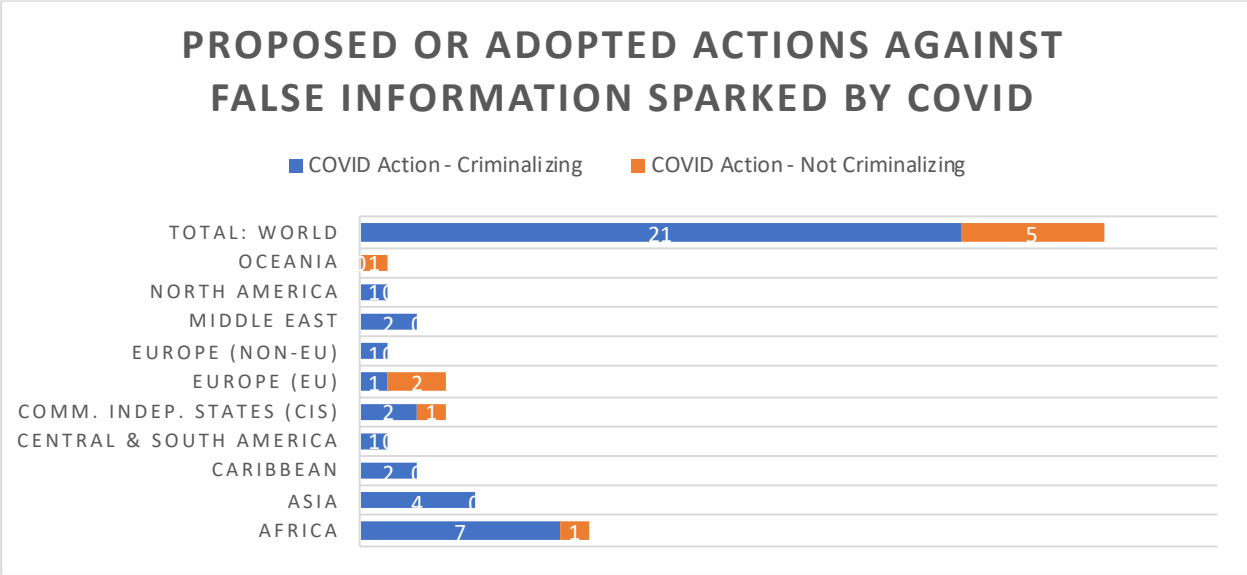
b. Global Trends



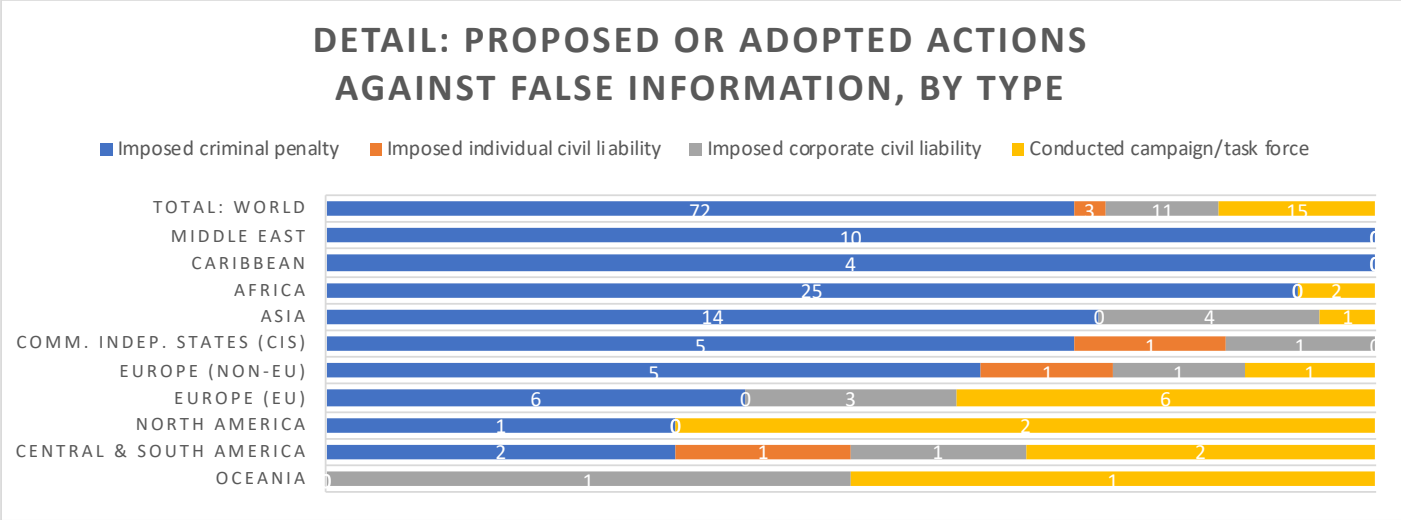
A majority of countries (52%) has considered or taken action against the spread of false information, and a majority of those countries (71%) has included criminal penalties. However, these countries that have considered or taken action criminalizing the spread of false information are in the minority overall, constituting 37% of countries. 11% of the criminalization that has occurred or is considered utilizes anti-defamation laws to achieve those ends.



The COVID-19 pandemic has sparked about a quarter of all actions proposed or taken against false information, and the majority of those pandemic-related actions have included criminal penalties. 24% of all the actions considered or taken against misinformation were in direct response to the COVID-19 pandemic, and 88% of COVID-19-prompted actions including the criminalization of false information. As of today, there is no telling how long these laws will be applied and if these laws criminalizing misinformation will cease to be applied once the pandemic is over.



As to non-criminal action, 30 countries have taken or are considering taking action other than criminalization: 3 countries imposing civil liability on the individual spreading false information; 11 countries imposing intermediary liability on the platform; and at least 16 countries setting up task forces or conducting news-correction campaigns.



For the purposes of analyzing trends, the following metrics were assessed for each of 10 regions based on standard deviation from the mean.

(i) Typical percentage of countries that have taken or considered action of any kind against false information.

It is somewhat common for countries to take action against misinformation; however, it is not common for every country or no country in a region to take action against misinformation. Based on the standard deviation from the mean, a statistically typical region fell in a range of 27-77% of countries taking some kind of action against false information.

(ii) Typical percentages of countries whose actions included criminalization of the spread of misinformation, disinformation, or other synonymous activities.

It is very common, for countries which consider or take action against misinformation or disinformation, to have that action include criminal penalties for spreading false information. Based on the standard deviation from the mean, in a statistically typical region, out of those countries who considered or took legislative action against false information, for 38-100% of those action-taking countries to considered or impose criminal liability.

(ii) Typical percentages of countries whose actions were taken in response to the COVID-19 pandemic.

It is somewhat uncommon for a country's action against misinformation to have be taken in response to the COVID-19 pandemic. However, having taken no action in response to the pandemic is also uncommon. Based on the standard deviation from the mean, a statistically typical region fell in a range of 10-38% of countries having taken or considered action in response to the COVID-19 pandemic.

(ii) Typical percentages of countries whose actions in response to the COVID-19 pandemic include criminalization.

It is quite common for actions taken against false information related to the COVID-19 pandemic to include criminalization of the act. Based on the standard deviation from the mean, a statistically typical region fell in a range of 53-100% of countries taking action in response to COVID-19 having done so via criminalization.

(5) Typical percentages of countries that prosecute misinformation through anti-defamation laws.

It is uncommon for criminalization of misinformation to be conducted through the utilization of anti-defamation laws. Based on the standard deviation from the mean, a statistically typical region fell in a range of 0-29% countries criminalizing misinformation to do via defamation laws.

c. Regional Trends

The obligations and liabilities of individuals, social media, and news platforms vary according to the circumstances of their geographic and digital environments. See the appendix for a more detailed breakdown regional data.

However, in the ten geographic regions studied, most metrics yielded results that were statistically typical, meaning they were within a standard deviation from the mean of countries where each action is taken. Africa and the Middle East were the only regions where the five aforementioned metrics of percentage of countries taking action against the spread of false information, etc. were consistently statistically typical relative to global trends.

Africa's statistical typicality of action may be in part due to the number of data points collected; African countries represented a plurality of data in this study. While it is theoretically possible to re-assess the global trends by removing Africa from the data pool, that would not be suitable for the purposes of this study. All data from the Middle East, a comparatively smaller region, also turned out statistically typical for the region, indicating Africa is not a sole outlier from the rest of the regions assessed. Additionally, the significant influence African countries have on global trends is not unique to this study; the impact African countries have on the data collection in this experiment is roughly similar to the level of impact this region has in voting in the United Nations General Assembly. In short, it would not be suitable to ignore trends in Africa on the basis that the size of the region gives it undue influence in data analysis. While the actions of countries in Africa and the Middle East can be considered statistically typical relative to global decision-making around misinformation laws, there were instances in the other eight regions studied that indicate statistically atypical behavior.

In Asia, 79% of countries surveyed have taken or considered taking action against false information. In this metric, Asia falls further beyond the mean than one standard deviation,

indicating it is significantly more common in Asia than in other regions for a nation to take or consider action against false information.

Among Caribbean nations, 50% of the nations who have taken or considered taking action against false information have done so in response to the recent COVID-19 pandemic. This significantly high percentage indicates that Caribbean nations have prioritized preventing false information in response to the pandemic over information that impacts national security, public order, elections, or other public matters. Additionally, 50% of the persecution of false information in Caribbean nations was done through the use of anti-defamation laws, which is a significantly higher percentage than in other regions. In these legal systems, the line between spreading false information and defamation is blurred.

In Central and South America, 33% of countries surveyed have criminalized or considered laws criminalizing the spread of false information. This indicates, relative to the rest of the world, a significantly low tendency among Central and South American countries to pass or propose laws that criminalize the spread of false information.

Among members of the Commonwealth of Independent States, 43% of the actions taken against false information were in response to the COVID-19 pandemic. As with the Caribbean states, these nations have ostensibly prioritized preventing false information in response to the pandemic over fighting information that impacts other public matters.

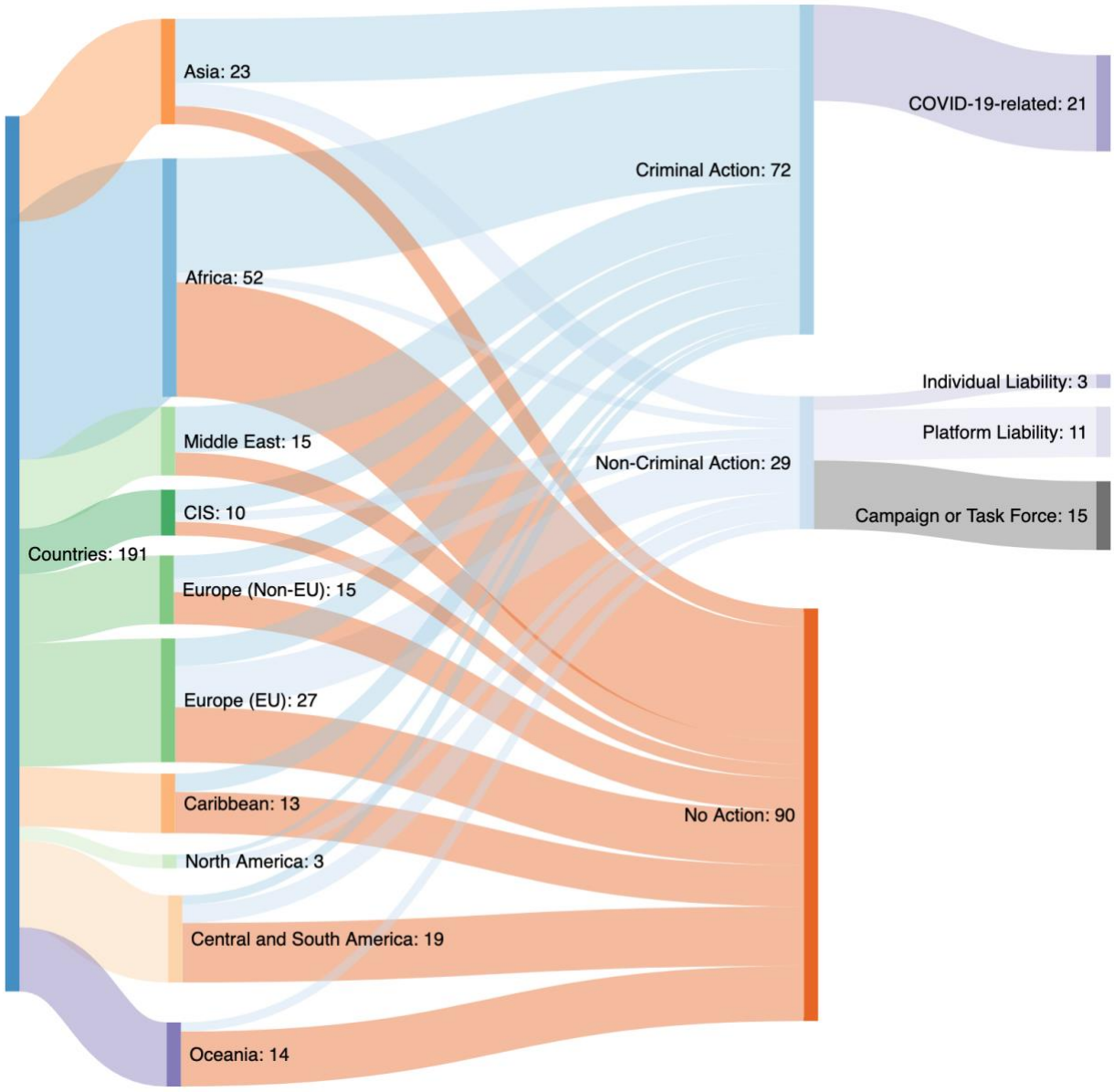
Among European countries that are not members of the European Union, there tended to be a significantly high incidence of countries utilizing anti-defamation laws for the purposes of prosecuting misinformation. 40% of the actions taken or considered to criminalize false information by non-EU European countries leaned on anti-defamation law.

Among EU members, there tended to be a significantly low incidence of countries criminalizing the spread of false information in response to the COVID-19 pandemic. While these countries were statistically typical in the fact that they enacted or considered false information laws in response to the pandemic, it seems these countries are generally more likely to take non-criminal action than criminal action; this trend was similarly reflected in their nearly statistically atypical tendency to favor non-criminal actions, such as intermediary liability, campaigns or task forces (60% of actions taken).

The North American group consists of only three countries, Canada, Mexico, and the United States. All three took or considered some action against false information. One included criminal penalties, while the other two did not. However, the region of three nations does not carry enough statistical weight to display trends that accurately represent their government practices.

In Oceania, a significantly low 14% nations have taken or considered taking any action against misinformation. Additionally, it is statistically atypical that no nations in Oceania have criminalized or considered criminalizing the spreading of misinformation.

Below is a diagram that represents how regions are addressing false information. Note that three nations (Republic of Congo, Niger, and North Korea) could not be represented due to a lack of available information regarding their national laws. To interpret the diagram, begin at the left-hand edge and visually track the proportion of countries in each region that proposed or took each type of action: criminal penalties, non-criminal response, or no action at all.



2. Qualitative Analysis

a. Trends in the Content of Laws Criminalizing False Information

Among the countries analyzed, the most common trend seemed to be that law enforcement decisions to charge and prosecute people found to have spread false information are contingent on the impact of the dissemination. Laws criminalizing the spread of false information

tend to focus on the consequences of the action rather than on the action itself. It seems that countries primarily seek to prosecute those whose spread of false information (or misinformation, disinformation, rumors, propaganda, hoax news, or fake news depending on the jurisdiction) led to a harm of a person's reputation, a threat to national security, a detriment to public health, or a diminution of public trust in government.

Although the term "disinformation" was not used in statutes, laws tended to describe punishable offenses as the spread of false information intentionally with the goal of convincing someone to act on the information as if it were true. Most countries do not specify whether their laws could be applied to cases of inadvertent publication of misinformation, with some intentionally limiting their scope to disinformation and others showing evidence of using disinformation laws to prosecute cases of misinformation. Those instances have been noted below.

The Library of Congress, which also created an index of initiatives in various countries combating false information, similarly concluded that "there is no common position regarding the definition of 'fake news' and its scope."[659] The Library of Congress took its analysis a step further by noting that "broad definitions" of misinformation crimes are usually found in the laws of "countries that are rated low in indices related to freedom of speech." "Such laws," according to the Library of Congress, are "often viewed by human rights organizations as government attempts to further restrict free speech and stifle opposition."

b. Trends in National Courts' Responses to These Criminal Laws

Often, laws criminalizing the spread of false information are challenged before national courts on the grounds that they violate the right to freedom of expression as guaranteed in national constitutions. National courts have considered whether the concept of "false information" is too vague to be used in a criminal statute; their inquiries have produced mixed outcomes. Kenya recently found that using such concepts did not violate free expression. In contrast, a 20-year-old Zimbabwean case and a more recent Zambian case found that using these concepts to criminalize speech did violate free expression. Additionally, in a landmark 1992 decision, the Canadian

[659] Peter Roudik, *Initiatives to Counter Fake News: Comparative Summary.* Summary, Washington, D.C.: Library of Congress, 2019.

Supreme Court declared a criminal law punishing false information unconstitutional due to vagueness.

The region where courts have addressed the matter in highest numbers is Africa. However, courts there do not always distinguish between defamation crimes and false information crimes. They sometimes adapt and apply defamation doctrine when deciding if false information crimes are compliant with the right of freedom of expression. Dicta in defamation cases often suggests that the holding would also apply to false information. It is hard to appraise if this entanglement in Africa between classical defamation doctrine and incipient false information doctrine is just rhetorical, or if both doctrines do share the same principles.

3. Selected Countries by Region: Criminal Laws and Court Cases on their Constitutionality

Below is a selection of countries, organized according to region, that have passed laws criminalizing the spread of false information. Each entry includes a description of the relevant law (including whether it is focused on disinformation, negative externalities, both, or something else) and, where appropriate, the implementation and jurisprudence surrounding the law.

a. Africa

Burkina Faso and Egypt have both criminalized the spread of false information with lengthy prison sentences. South Africa has taken a more targeted and measured approach, lightly criminalizing with prison sentences of up to 6 months disinformation related to COVID-19.

Kenya, Tanzania, Zambia, and Zimbabwe have all passed laws criminalizing the spread of false information that have been held, by international or national courts, to violate free expression. However, other such laws remain on the books in Zimbabwe, and Kenya's high court has upheld another of its criminal laws targeting the spread of false information, despite its authorization of lengthy prison sentences.

(i) Burkina Faso

In June of 2019, the National Assembly of Burkina Faso criminalized the dissemination of information that could undermine public order or security operations.[660] The national penal code defines this offense as intentionally communicating or publishing "false information of a kind to suggest that a destruction of property or an attack against persons has been committed or will be committed." A prosecutor or "any person having interest" may petition a judge to force the removal of the false information. The publisher faces prison time of up to 5 years and a fine of 10 million Central African francs ($17,000). The law's element of intentionality suggests it is primarily concerned with disinformation.

(ii) Egypt

Since 2003, Egypt has criminalized spreading "fake news." Specifically, the Egyptian penal code criminalizes the offense of deliberately spreading abroad false information or rumors about the country's internal conditions with the aim of weakening confidence in the country's economy or undermining its prestige. This offense is punishable by imprisonment for between 6 months and 5 years.[661] Egyptian authorities have arrested journalists and social media users accused of publishing false information.[662]

Additionally, under Article 19 of Egypt's Press and Information Regulation law and the Supreme Council for Media Regulation, the Egyptian government regulates social media accounts with followings of 5,000 or more accounts. Any account on social media platforms, such as Facebook or Twitter, or blog that has more than 5,000 followers will be treated like a media outlet, whose website can be blocked by the Supreme Media Council for publishing "fake news." Media

[660] Burkina Faso Penal Code of 2018, Book III, Title I, Article 312-13. *See* Nicolas Boring, "Burkina Faso: Parliament Amends Penal Code," Global Legal Monitor Entry, Sept. 20, 2019, Washington, D.C.: Library of Congress. Available at: https://www.loc.gov/law/foreign-news/article/burkina-faso-parliament-amends-penal-code/

[661] Egyptian Penal Code, Article 80(d), Law No. 58 of 1937, *as amended by* Law No. 95 of 2003, Vol. 25, *Al Jaridah Al-Rasmyiah*, June 19, 2003. *See* George Sadeck, "Initiatives to Counter Fake News: Egypt," Global Legal Monitor Entry, Apr. 2019, Washington, D.C.: Library of Congress. Available at: https://www.loc.gov/law/help/fake-news/egypt.php

[662] The Quarterly Report on the State of Freedom of Expression in Egypt 3rd Quarter (July-Sept. 2018), Freedom of Thought and Expression (Oct. 16, 2018). Available at: https://afteegypt.org/en/afte_releases/2018/10/16/16036-afteegypt.html, archived at https://perma.cc/AR53-CC8N; The Quarterly Report on the State of Freedom of Expression in Egypt 2nd Quarter (Apr.-June 2018), Freedom of Thought and Expression (July 10, 2018). Available at https://afteegypt.org/en/afte_releases/2018/07/10/15525-afteegypt.html, archived at https://perma.cc/LU79-VK8M.

outlet directors or website administrators that refuse to "correct" false information within 3 days are subject to hefty fines (between approximately 2,855 and 5,711 USD .[663]

(iii) Kenya

Kenya has passed a number of criminal laws targeting the spread of false information, some of which impose lengthy prison sentences as penalties. Some have been upheld as in compliance with constitutional free expression, while others have not.

Some provisions establish 'lightly' criminalized offenses with statutory penalties of up to 3 months imprisonment. One provision penalized using a telecommunications system to "send a message that he knows to be false" intended to cause annoyance or inconvenience to someone else with up to 3 months in prison.[664] However, the High Court of Kenya at Nairobi struck down this provision as unconstitutional, holding that its overbroad and vague terms violated free expression.[665] A second Kenyan penal code provision authorizes up to 2 years imprisonment for publishing "alarming publications," defined as publishing false information likely to alarm the public without having taken reasonable measures to verify its accuracy.[666]

Other Kenyan laws go farther, authorizing more than 2 years of imprisonment. For instance, in 2018, the President of Kenya signed into law the Computer Misuse and Cybercrimes Bill.[667] Under this paper's terminology, the law would likely qualify as targeting disinformation. The law establishes two criminal offenses related to false information. The first provision, Section 22, more lightly criminalizes publishing "false, misleading or fictitious data" or otherwise "misinform[ing]" with the intention of causing people to act on the information as if it is

[663] George Sadeck, "Initiatives to Counter Fake News: Egypt."
[664] Penal Code of 1930, § 66, 16 Laws of Kenya, Cap. 63; Kenya Information and Communications Act No. 2 of 1998, § 29. *See* Hanibal Goitom, "Initiatives to Counter Fake News: Kenya," Global Legal Monitor Entry, Apr. 2019, Washington, D.C.: Library of Congress. Available at: https://www.loc.gov/law/help/fake-news/kenya.php
[665] Robert Mabera, *Andare v. Attorney General* (High Court of Kenya, Apr. 19, 2016), at ¶ 35. Available at: http://kenyalaw.org/caselaw/cases/view/121033/.
[666] Penal Code of 1930, § 66, 16 Laws of Kenya, Cap. 63; Kenya Information and Communications Act No. 2 of 1998, § 29. *See* Hanibal Goitom, "Initiatives to Counter Fake News: Kenya."
[667] *President Kenyatta Signs Into Law The Computer Misuse And Cybercrimes Bill, 2018,* Press Release, Nairobi: Office of the President, 2018.

authentic.[668] Perhaps as a caveat, the law goes on to explain that under the Kenyan Constitution, the government may limit a person's freedom of expression only if he intentionally publishes misinformation that is likely to propagate war or incite violence, constitutes hate speech, advocates discriminatory hatred, or negatively affects others' reputations. The penalty for this action is a fine of 5 million shillings (50,000 USD), up to 2 years in prison, or both. Under the more severe second provision, Section 23, knowingly publishing false information "that is calculated [to] or results in panic, chaos, or violence" among Kenyans or even just "is likely to discredit the reputation of a person," is criminalized with a statutory penalty of up to 10 years in prison.

Shortly after the law was passed, the Bloggers Association of Kenya (BAKE), joined by Article 19 and other NGOs, challenged the constitutionality of both of the above-mentioned provisions[669] before Kenya's High Court.[670] The constitutionality of the law was challenged facially, rather than being tied to a specific application.

On February 20, 2020, the High Court held that provisions were in compliance with the right of freedom of expression as enshrined in the Kenyan Constitution. By doing so, the court came to opposite conclusions than it had in previous defamation and false information decisions: *Geoffrey Andare vs Attorney General and 2 others (2016)* and *Jackqueline Okula & another vs Attorney General & 2 others (2017)*. In *Andare*, the challenged provision criminalizing the publication of messages that the sender "knows to be false for the purpose of causing annoyance, inconvenience or needless anxiety to another person" was ruled unconstitutional due to its overbreadth and vagueness.[671] In *Okula*, the court struck down a criminal defamation provision as unconstitutional, declaring that "the offence of criminal defamation is not reasonably justifiable in a democratic society." [672]

[668] "Computer Misuse and Cybercrimes Act, 2018," Nairobi: Office of the President. Accessed June 10, 2020. *See* Hanibal Goitom, "Initiatives to Counter Fake News: Kenya."
[669] Sections 22 and 23 of the Act.
[670] *See Bloggers Association of Kenya (BAKE) v. Attorney General & 3 others* (High Court of Kenya 2020). Available at: http://kenyalaw.org/caselaw/cases/view/191276/
[671] *Andare v. Attorney General* (High Court of Kenya, Apr. 19, 2016), at ¶ 35. Available at: http://kenyalaw.org/caselaw/cases/view/121033/.
[672] *Jackqueline Okula & another v. Attorney General & 2 others* (High Court of Kenya, 2017). Available at: http://kenyalaw.org/caselaw/cases/view/130781/

As to vagueness and overbreadth, the *BAKE* court dismissed concerns that the term "false" was overbroad by arguing that it is a plain English word that does not require a legal definition.[673] The Court declares itself unconvinced that the "truth is not [a] necessary condition to the freedom of expression."[674] The High Court distinguished its past decisions of *Andare* and *Okula* by pointing out that provisions challenged in BAKE clearly "spell[ed] out both the actus reus and the *mens rea*" of the offense. To this outside observer, the level of specificity in the provisions at issue in *Andare* are similar to those in *BAKE*. Therefore, it could be argued that the Court changed its doctrine, criteria, or predisposition without explicitly announcing it.

As to necessity, the *BAKE* court dismissed the allegation that Section 22 (on the publication of false data) was not necessary by saying that the petitioner did not demonstrate that an alternative measure could be taken to realize Section 22's objective. With respect to Section 23, the court found it necessary to establish a law to control spreading false information which could threaten national security, pointing to both the post-election violence that broke out in Kenya during 2007 and 2008, and how swiftly and "often irredeemably" false information spreads on the Internet.[675]

As to the second part of Section 23, which also criminalizes spreading false information likely to discredit another's reputation, the court drew parallels to the criminalization of defamation. The court cited a 2008 case from South Africa for the proposition that a civil remedy does not negate the need for a criminal remedy.[676] In a previous case, *Jackqueline Okula & another vs Attorney General & 2 others* (2017), the High Court of Kenya had struck down a criminal defamation provision as unconstitutional, declaring that "the offence of criminal defamation is not reasonably justifiable in a democratic society."[677] However, in BAKE, the court found that cyber libel could be criminalized even if offline libel could not, reasoning that an online publisher of libel had a greater "ability to evade identification" and "reach a far wider audience" thereby causing greater harm, when compared to offline publishers.

[673] *Bloggers Association of Kenya (BAKE)*, at ¶ 50.
[674] *Id.*, at ¶ 61.
[675] *Id.*, at ¶¶ 44–45.
[676] *Id.*, at ¶ 53. *See Hoho v. The State* (Supreme Court of Appeal of South Africa, 2008).
[677] *Jackqueline Okula & another v. Attorney General & 2 others*.

(iv) South Africa

In March of 2020, the South African government published new regulations making it a lightly criminalized offence to publish any statement on any platform with the goal of "deceiving" people about COVID-19, a person's infection status, or government action in response to the pandemic. Anyone found guilty of violating this law may face a fine, prison for up to 6 months, or both.[678] Emphasis on the deception of others indicates this law is focused on disinformation.

Tanzania

In 2016 ,the Parliament of the Republic of Tanzania enacted the Media Services Act which declares it a heavily criminalized offence to use a media service for the purpose of publishing "information which is intentionally or recklessly falsified" such that it threatens defense, public safety, public order, economy, morality, or public health of the nation, or is "injurious to the reputations, rights and freedom of other persons." This last provision is distinct from general defamation, which is addressed in a different part of the law. The law also criminalizes publishing information that is "maliciously or fraudulently fabricated," any statement that is "false or without reasonable grounds for believing it to be true," and information that is "likely to cause fear and alarm to the public or to disturb the public peace." The punishment for violating this law is a fine of 5 to 20 million shillings ($2,150 to $8,600), imprisonment for 3 to 6 years, or both. Emphasis on intentionality in this law indicates it is focused on combatting disinformation.

This law was eventually challenged before the East Africa Court of Justice, which declared some of its provisions to violate the right free expression. The case was briefed in the international law section of this report.

(v) Zambia

In the *Chipenzi v. The People* 2014 ruling, the High Court of Zambia found that prohibiting the publication of false information violated the Zambian Constitution because the state had no

678 *Government Gazette.* Notice, Pretoria: Department of Co-operative Governance and Traditional Affairs, Republic of South Africa, 2020.

reasonable justification for limiting the freedom of expression.[679] In 2013, journalists published an article claiming that Zambia's secret police had recruited several foreign militias into the police service. The government charged them under a law that punishes the dissemination of false information that is "likely to cause fear and alarm to the public or to disturb the public peace."

The main issue on appeal to the High Court of Zambia was whether or not it was reasonable to limit freedom of expression by criminalizing false information likely to cause public fear. The court found the prohibition was overly broad, finding that the law circularly restricts "any statement which does not meet the majority definition of truth," and is thus susceptible to abuse by prosecuting " news which is unpopular in the ears of those in authority." The court also rejected the idea that the law helped protect public order because it found that prosecution and conviction were not dependent upon any public fear or disturbance actually occurring. According to the court, the law in its current form protected against remote dangers from the spread of false news that could not properly and explicitly be quantified, and thus violated free expression.

(vi) Zimbabwe

In 2000, in the case of *Chavunduka and others v Minister of Home Affairs,* the Supreme Court of Zimbabwe declared a 1961 criminal law on "false news" to be unconstitutional on the basis that the law was not precise enough to allow people to regulate their conduct, and that the offence of "false news" was unnecessary, overbroad, and not "reasonably justifiable in a democratic society."[680] However, as recently 2020 the Zimbabwean government announced an intent to rely on a seemingly equally vague penal code provision criminalizing "false statements prejudicial to the State" in connection with a COVID-19 necessitated lockdown.

The *Chavunduka* case arose when, in 1999 the editor and senior journalist of the *Standard* newspaper in Zimbabwe were arrested and charged with publishing "false statement likely to cause

[679] *Chipenzi v. The People* (High Court for Zambia, March 2014). Available at: https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2015/03/Chipenzi-v.-The-People-HPR032014.pdf
[680] *Ibid*, at 25, 26

fear, alarm or despondency" under section 50 (2)(a) of the Law and Order (Maintenance) Act. They had published an article describing a failed coup d'etat.[681]

The Supreme Court of Zimbabwe determined first that the provision was too vague to pass the legality test under the Zimbabwean constitution. Because the provision was concerned with likelihood rather than reality, it was susceptible to too wide an interpretation. As a result, it had an unacceptable "chilling effect" on freedom of expression.[682] Another source of uncertainty was the overbreadth of "fear, alarm or despondency," as the court argued that any newsworthy event is likely to cause these emotions.[683]

The court went on to find that even if the terminology was not deemed overly vague, the law would not be constitutional, as it would fail the other prongs of a three-part test that parallels international law.[684] Under this test, limitations on free expression must be (a) authorized by law (declared equivalent to "prescribed by law" as used in international human rights treaties), (b) enacted in the interests of a legitimate aim such as public safety or public order, and (c) "reasonably justifiable in a democratic society." The court found the provision not necessary, or—in this application—proportionate to attain a legitimate aim.

Nevertheless, as part of a public health directive ordering a national lockdown in response to the COVID-19 pandemic, in 2020 the Zimbabwean government announced that anyone who publishes "false news" about any "public officer" involved with enforcing or the lockdown or about any "private individual" that has the effect of "prejudicing the State's enforcement of the national lockdown" may be prosecuted under a Zimbabwe Criminal Law Code Section 31. Section 31 prohibits "publishing or communicating false statements prejudicial to the State." Violators may be fined, sentenced to prison for up to 20 years, or both.[685] While the law stipulates there the dissemination of false information must have a detrimental effect, it is unclear whether even

681 *Chavunduka and others v. Minister of Home Affairs and another* [2000] JOL 6540 (ZS), at 4 (Supreme Court of Zimbabwe 2000). Available at https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2017/08/Chavunduka-v-Minister-of-Home-Affairs-Zimbabwe9610.pdf.
682 *Ibid*, at 14.
683 *Ibid.*
684 *Ibid*, at 19, 26
685 "Public Health (COVID-19 Prevention, Containment and Treatment) (National Lockdown) Order, 2020," *Harare: Zimbabwean Government Gazette Extraordinary*, Republic of Zimbabwe, 2020.

inadvertent sharing of such misinformation may be prosecuted. It is also unclear how the law complies with the holding of the 2000 *Chavunduka* case.

b. Asia

Bangladesh, China, and Singapore all have laws targeting misinformation or disinformation which authorize penalties of more than 2 years in prison. India uses various criminal codes for offenses, such as sedition, to prosecute the spread of "fake news." One such criminal provision was struck down as a violation of free expression by the Indian Supreme Court. Malaysia passed a law criminalizing the spread of "fake news" in 2018 but repealed it in 2019.

(i) Bangladesh

Bangladesh has passed two laws which criminalize the spread of false information on certain topics or with the potential to have specified negative consequences. Both prescribe lengthy prison sentences, and one authorizes a life sentence of imprisonment for repeated violations.

Bangladesh passed the 2006 Information and Communication Technology (ICT) Act to support the growth of e-commerce.[686] However, Article 57 of the ICT Act enables the prosecution of anyone who deliberately digitally publishes material that is "fake" and "creates [the] possibility" of the following negative consequences: "deteriorat[ing] law and order," prejudicing the state's image, "hurt[ing] religious belief" or "instigat[ing] against any person or organization." The wording of this law may enable prosecution of misinformation as well as disinformation. This section authorizes penalties of between 7 and 14 years in prison.

Between 2013 and 2018, Bangladeshi police submitted 1,271 charge sheets under the ICT Act, with the majority being under Article 57. [687] Notable prosecutions include that of Nasiruddin Elan and Adilur Rahman Khan, officials of "prominent human rights organization Odhikar."[688]

[686] Md Nayem Alimul Hyder, "ICT Act 2006: Making e-Commerce Effective," *The Financial Express*, Jan. 15, 2017. Available at: https://thefinancialexpress.com.bd/views/ict-act-2006-making-e-commerce-effective.
[687] David Bergman, "No Place for Criticism: Bangladesh Crackdown on Social Media Commentary," *Human Rights Watch*, July 2, 2018. Available at: http://www.hrw.org/report/2018/05/09/no-place-criticism/bangladesh-crackdown-social-media-commentary.
[688] Sam Zarifi, et al., "Bangladesh: Information and Communication Technology Act Draconian Assault on Free Expression," International Commission of Jurists, Nov. 20, 2013.

The two were accused of intentionally misreporting the number of people killed during a protest that was broken up by police in 2013. Elan and Khan were found guilty, and a 2017 appeal of their case to the High Court of Bangladesh was rejected.[689] Other journalists and bloggers have also been detained under the ICT. One such blogger was detained for releasing a transcript of a conversation between legal experts questioning the independence of the nation's domestic war crimes tribunal, likely on the theory that such claims may cause a deterioration in law and order or prejudice the state's image, as discussed above. Four others have been arrested for derogatory commentary about Islam, which pertains to the ban on material that may cause "hurt to religious beliefs."

In 2018, Bangladesh passed the Digital Security Act, which provides for the criminalization of using digital devices to spread "negative propaganda" regarding two wars, as well as the publication of "intimidating" or "distorted" information.[690] The Act was passed unanimously by the Parliament of Bangladesh "to ensure digital security and prevent crimes committed on digital platforms."[691] The Digital Security Act, which expands on the ICT law, criminalizes "spreading propaganda" that opposes Bangladesh's national anthem, flag, or independence efforts of the 1970s Liberation War. The maximum sentence for spreading this propaganda is 10 years. However, the act authorizes penalizing repeated violation of this law with a life sentence.

Shortly after the Digital Security Act passed, another bill, the Broadcast Act was proposed to strengthen media laws.[692] The law is still actively being considered by the Bangladesh government. This act would pertain to all media, not just digital media, criminalizing "the telecasting, broadcasting, or publishing of any statement deemed to be against the country or against public interest; sharing any misleading or untrue information or data on a talk show; and broadcasting any show or advertisement contrary to national culture, heritage and spirits." The Broadcast Act also establishes a commission that could issue fines to or revoke licenses from

689 Bergman, "No Place for Criticism: Bangladesh Crackdown."
690 *Ibid.*
691 "Bangladesh: Digital Security Act 2018," *Article 19*, Nov. 2019. Available at: www.article19.org/wp-content/uploads/2019/11/Bangladesh-Cyber-Security-act-2018-analysis-FINAL.pdf.
692 Euan Rocha, "Factbox: Bangladesh's Broad Media Laws," *Reuters*, Dec. 12, 2018.

outlets violating the law as well as recommending the prosecution of individuals, who could be imprisoned for up to 7 years if found guilty.

(ii) China

In 2015, Section 29a1 of the Criminal Law of the People's Republic of China was amended to penalize anyone who "fabricates a false dangerous situation, an epidemic situation, a disaster situation or a warning situation and spreads it on an information network or other media, or knowingly spreads it on an information network or other media and seriously disturbs public order."[693] The wording of this law indicates it may be used to prosecute both misinformation and disinformation. An individual convicted under this law could face 3 to 7 years in prison.

In 2016, the National People's Congress Standing Committee adopted the PRC Cybersecurity Law. This law prohibits manufacturing or spreading fake news online that disturbs the country's economic and social order.[694] Unavailability of the law in English precludes further analysis.

(iii) India

Indian authorities have used various criminal codes to prosecute people found to have spread false information, although these provisions do not require that the information be false. One such provision was held to violate free expression by the Indian Supreme Court. Most recently, Indian authorities have used these laws to make arrests of individuals accused of spreading false information about COVID-19 and the government's response to the pandemic.[695]

Indian Penal Code provisions used for these purposes include Section 124A on sedition, originally drafted by British colonial authorities, which makes it a criminal offense to "bring[ ] or attempt[ ] to bring into hatred or contempt, or excites or attempts to excite disaffection towards,

693 Laney Zhang, "China: Initiatives to Counter Fake News," Library of Congress, Apr. 2019. Available at: www.loc.gov/law/help/fake-news/china.php.
694 Zhang, Library of Congress. *See* PRC Cybersecurity Law (adopted by the NPC Standing Committee on Nov. 7, 2016, effective June 1, 2017). Available at: http://www.npc.gov.cn/npc/ xinwen/2016-11/07/content_2001605.htm (in Chinese), *archived at* https://perma.cc/3HAP-D6MZ.
695 Bhavya Dore, "Fake News, Real Arrests," *Foreign Policy*, Apr. 17, 2020.

the Government established by law in India."[696] Violation is punishable by various combinations of penalties, including one for 3 years in prison.[697] Other criminal provisions in India prohibit making statements that promote enmity against a group, insult with intent to provoke violence, or are conducive to "public mischief," in addition to blasphemy and defamation laws, all punishable by up to 2 or 3 years in prison. Other than defamation, these laws do not have elements or defenses related to truth or falsity. In contrast, the criminal offense of making a "statement purporting to be a statement of fact" does require that the statement be false and the speaker believe it to be false, but punishes violation with only a fine.[698]

Until the law was struck down by the Indian Supreme Court, India also used the 2000 Information Technology Act, which banned "dissemination of information by means of a computer resource or a communication device intended to cause annoyance, inconvenience or insult" and authorized up to 3 years in prison as a statutory penalty[699] This provision in the law was overturned in 2015 after the Indian Supreme Court ruled that a prosecution under the law violated free expression.[700] The case involved a man who posted a falsified recording pertaining to a terrorism incident.

(iv) Malaysia

In 2018 the Malaysian government passed the Anti-Fake News Act, which criminalized the spread of fake news with lengthy prison sentences. However, a new parliamentary majority whose leader had promised to repeal the law won an election repealed the act in 2019.[701]

[696] Tariq Ahmad, "Government Responses to Disinformation on Social Media Platforms: India," Global Legal Monitor Entry, Library of Congress, Sept. 2019. Available at: https://www.loc.gov/law/help/social-media-disinformation/india.php
[697] Indian Penal Code, No. 45 of 1860. Available at: https://perma.cc/49VP-ZC6C.
[698] Indian Penal Code, Section 171G. Available at: https://perma.cc/49VP-ZC6C.
[699] Ahmad, "Government Responses to Disinformation on Social Media Platforms: India."
[700] *Shreya Singhal v. Union of India*, Writ Petition (Crim.) No. 167 of 2012 (Indian Supreme Court, Mar. 24, 2015). Available at: https://perma.cc/X8QZ-7TV8
[701] Anti-Fake News (Repeal) Act, Kuala Lumpur: Parliament of Malaysia. Rozanna Latiff, "Malaysia parliament scraps law penalizing fake news," *Reuters*, Oct. 9, 2019. Available at: https://www.reuters.com/article/us-malaysia-politics-fakenews/malaysia-parliament-scraps-law-penalizing-fake-news-idUSKBN1WO1H6

Before repeal, the 2018 law made it an offence to maliciously create or publish "fake news" and to refuse to remove such news when mandated to do so by a court.[702] The provision authorized penalties of up to 6 years in prison. The act also specifically provided for extraterritorial application, even by non-Malaysians, as long as the fake news related to Malaysia or a Malaysian citizen. This law did not require the offending content have negative externalities, only that it be "wholly or partly false." However, the illustrations of offending content specifically precluded prosecuting individuals who inadvertently published information that they were made to believe was true. That, in addition to the required *mens rea* of malice, suggests that the law seemed to be concerned only with disinformation.

(v) Singapore

In 2019, Singapore passed and implemented the Protection from Online Falsehoods and Manipulation Law, which broadly prohibits "communication of false statements of fact" likely to have specified negative consequences, using "bots" (defined as an automated computer program) to communicate these false statements, and providing services that communicate these false statements "for the purpose of mis-leading end-users in Singapore."[703] The law also authorizes governmental officials to direct individuals, platforms, and internet service providers to correct false information or prevent its dissemination.

The law targets disinformation rather than merely misinformation. To justify prosecution under the provision on communicating false statements, the defendant must have known or had reason to believe that his communication was not only a false statement of fact but also likely to be prejudicial to the security of any part of Singapore, its public health, public safety, public tranquility, or public finances, influence the outcome of a national election, incite enmity between groups, or diminish public confidence in the government. A "statement of fact" is defined in the law as a statement which a reasonable person might interpret as a fact. A "false statement of fact" is completely or partially "false or misleading." Violation of this provision by individuals or a group may be penalized with a fine of up to 500,000 Singaporean dollars (350,000 USD), up to 5

[702] *Anti-Fake News Act*, Kuala Lumpur: Parliament of Malaysia, 2018.
[703] *Protection From Online Falsehoods and Manipulation Act*, Parliament of Singapore, Apr. 2019.

years of imprisonment, or both. Use of "bots" to aid the spread of false statements carries a more severe punishment: up to 10 years in prison.

c. Central and South America

(i) Brazil

Several pieces of legislation have been proposed in the National Congress of Brazil, all of which pertain to punishing the spread of "fake news."[704] They are concerned primarily with the effect of the false information on matters of "public interest" or if the false information incites violence.[705] It is uncertain whether these laws focus solely on disinformation or whether they could be applied to inadvertent spreading of misinformation as well. None of these proposed laws seem to have passed thus far.

d. Commonwealth of Independent States

(i) Russia

In 2020, the Russian Federation amended the section of its criminal code that previously made it a crime to report false information on terrorist activity to add two new disinformation offenses.[706] English translations of the amendments are not yet available, but the Kremlin reports that the first new provision, Article 2071, prohibits "public distribution of deliberately misleading information about circumstances that pose a threat to the life and safety of citizens." The second, Article 2072, "establishes liability for public distribution of deliberately misleading socially important information that resulted in grave consequences due to negligence."[707] The first provision carries statutory penalties of fines or up to 1 year in prison, while the second provides for fines or up to 3 years in prison.[708] The inclusion of the phrase "deliberately" in the Kremlin

[704] Gaspar Pisanu, "Brazil: 'fake news' proposals add uncertainty to institutional crisis," AccessNow, Apr. 27, 2018.
[705] Amends Decree-Law No. 2,848, of Dec. 7, 1940 - Penal Code, Federative Republic of Brazil: Chamber of Deputies, March 27, 2018.
[706] *See* International Press Institute, "New 'fake news' law stifles independent reporting in Russia on COVID-19," International Press Institute, May 8, 2020.
[707] *Amendments to Russian Criminal and Criminal Procedure Codes*, Moscow: Office of the President, Russian Federation. Apr. 1, 2020.
[708] Federal Law No. 100-FZ, "On Amendments to the Criminal Code of the Russian Federation and Articles 31 and 151 of the Criminal Procedure Code of the Russian Federation." Available at: http://publication.pravo.gov.ru/Document/View/0001202004010073.

descriptions indicates that these amendments focus on disinformation, but in practice, the provisions seem to have also resulted in fines in cases of misinformation.[709]

On September 2016, Russia's Supreme Court upheld a conviction for disseminating "knowingly false information about the activities of the USSR during the Second World War."[710] This decision is not strictly one of the misinformation cases that this report seeks to map, since it involves a so-called memory law. As was explained in the international law section of the report, the treatment of memory laws under international law is clearer, so the report does not delve deeply into state practice regarding these laws.

However, the decision evinces how comfortable Russia's Supreme Court is with the law criminalizing the spread of information branded by the executive branch as false, and it constitutes precedent that could be extrapolated to false news. The Supreme Court relied on the testimony of prosecution experts who concluded that the statements "did not correspond to the reality recognized on the international level."[711] However, the statements merely consisted in affirming that "the Communists…actively collaborated with Germany in dividing Europe according to the Molotov-Ribbentrop Pact,"[712] which is the version of history generally accepted as true in other countries.

e. Europe (EU Members)

France prohibits the bad-faith dissemination of false news but only authorizes fines as penalties. Italy has entertained but not passed a proposed law that would have criminalized the spread of false news with up to 1 year's imprisonment.

While individual European nations have taken action on their own, the European Commission (EC) has also published some literature relating to EU members' role in protecting their countries from disinformation, rather than misinformation. Both the EU Commission and the Council of Europe, in their independent reports, advocate against the use of such terms as "fake

709 Daria Litvinova, "Fake news or the truth? Russia cracks down on virus postings," Moscow: Associated Press, Apr. 1, 2020.
710 Gleb Bogush, "Russia's Supreme Court Rewrites History of the Second World War," Ejil: Talk, Oct. 28, 2016.
711 *Ibid.*
712 *Ibid.*

news," branding them as "woefully inadequate" and "misleading."[713] In March 2018, the European Commission published its Final Report of the High-Level Expert Group (HLEG) on Fake News and Online Disinformation. The Report reviewed best practices and suitable responses for fighting fake news and online disinformation and outlined self-regulatory standards of conduct.[714] The EC also warned that legislation could follow if the voluntary code does not lead to "measurable effects."[715] In the same report, the EC issued an Action Plan protecting the electoral process.[716] Countries throughout the EU can look to guidelines established in the report when drafting their laws surrounding disinformation.

(i) France

In 2018, France passed a law combatting disinformation. According to a description in English on a French government website, the law is focused on false information that aims to influence voters, and stipulates "the fake news must be manifest . . . disseminated deliberately on a massive scale, and lead to a disturbance of the peace or compromise the outcome of an election."[717] While the law authorizes courts to order "any proportional and necessary measure" to stop "deliberate, artificial or automatic and massive" online dissemination of false or misleading news, it does not authorize criminal prosecutions.[718]

Criminal prosecutions of disinformation are instead authorized in France by a 1881 law still in force which prohibits disturbing public peace via bad-faith dissemination of fake news and authorizes only fines, not imprisonment. The French Electoral Code outlaws disseminating false rumors that affect an election's results, punishable by fines or up to 1 year's imprisonment.

(ii) Italy

[713] "A Multi-Dimensional Approach to Disinformation Report of the Independent High Level Group on Fake News and Online Disinformation," European Commission, March 2018. Available at: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=50271.
Wardle, "Information Disorder."
[714] *Ibid.*
[715] *Ibid.*
[716] *Ibid.*
[717] "Against information manipulation," Paris: National Assembly, French Republic, Nov. 20, 2018. Available at: https://www.gouvernement.fr/en/against-information-manipulation
[718] Nicholas Boring, "Initiatives to Counter Fake News: France," Library of Congress, Apr. 2019. Available at: https://www.loc.gov/law/help/fake-news/france.php#_ftn11

Italy has not criminalized the spread of false news, though a bill proposed in 2017 would have done so with respect to disinformation.

In 2018, Italy created an online portal for citizens to report misinformation to law enforcement. The Postal Communications Police, a specific unit of the state police that investigates cybercrime, fact-checks the stories and, if it determines laws were broken, may pursue legal action. There is no legal definition of "fake news" that would be applied to these cases, although relevant laws might include fraud or disturbance of the public order.[719]

In 2017, a bill "to prevent the manipulation of online information, guarantee transparency on the web and encourage media literacy" was proposed in the Italian Senate.[720] The proposed legislation criminalizes digital publication or dissemination of "fake, exaggerated, or biased news" using information that is "manifestly unfounded or untrue." Minor offenses would be penalized with a fine, whereas offenses that are "clearly harmful to the public interest, or if it alarms the public" could be punished with up to 1 year's imprisonment. False information "aimed at undermining democracy in Italy or is part of a hate campaign against an individual" is penalized with a sentence of up to 2 years of imprisonment and a larger fine. Overall, the proposal seems to target disinformation rather than misinformation.

In 2015, Italy's antitrust authority fined TripAdvisor $550,450 after a national hotel association and the country's consumer protection agency accused the company of emphasizing "the authentic and genuine nature of reviews, persuading consumers to believe the information is always reliable and reflects real tourist experiences."[721] This accusation could match the definition of disinformation, but would also qualify as fraud since the alleged deception was spurred by a profit motive. A higher court later overturned the ruling, noting that TripAdvisor "explicitly acknowledged that it was not able to check the truthfulness of its reviews and advised users to

[719] Daniel Funke, "Italians Can Now Report Fake News to the Police. Here's Why That's Problematic," *Poynter*, Jan. 19, 2018. Available at: https://www.poynter.org/fact-checking/2018/italians-can-now-report-fake-news-to-the-police-heres-why-thats-problematic/

[720] Christina la Cour, "Governments Countering Disinformation: The Case of Italy," *Disinfo Portal*, Nov. 20, 2019. Available at: disinfoportal.org/governments-countering-disinformation-the-case-of-italy/

[721] "Tripadvisor Escapes Fine in Italy over Fake Reviews," *Bangkok Post*, July 14, 2015. Available at: www.bangkokpost.com/world/621288/tripadvisor-escapes-fine-in-italy-over-fake-reviews

consider broader trends in user-generated recommendations."[722] In its original ruling, the antitrust authority said TripAdvisor "failed to adopt controls to prevent false reviews." It gave TripAdvisor 90 days to remedy the shortcoming.[723]

f. Middle East

(i) Lebanon

There have been news reports of journalists being convicted of defamation and false-information offences for publishing stories about political leaders.[724] However, the unavailability of detailed information about these prosecutions and the justifying criminal provisions in English limits the reliability of these observations as to the existence of a fake-news criminal offense in Lebanon. These prosecutions may be conducted under the framework of Lebanese laws criminalizing defamation, incitement of enmity between groups, publishing materials "contrary to morality and public morals," or publishing insults to recognized religions that would disrupt the public peace, which often authorize 1, 2, or 3 years of imprisonment and fines.[725]

(ii) United Arab Emirates

In April, in an effort to restrict dissemination of false information relating to the COVID-19 pandemic, the United Arab Emirates issued a directive forbidding the publication of "medical information or guidance which is false, misleading or which hasn't been announced officially."[726] Violators will pay a fine of up to AED 20,000 ($5,500) for publishing "misleading" information.[727]

[722] Philip Willan, "TripAdvisor Gets Italian Fine Overturned," *CIO*, IDG News Service, July 15, 2015. Available at: www.cio.com/article/2948713/tripadvisor-gets-italian-fine-overturned.html.
[723] "Italy Fines TripAdvisor €500,000 over False Reviews," *The Guardian*, Dec. 23, 2014. Available at: www.theguardian.com/travel/2014/dec/23/italy-fines-tripadvisor-500000.
[724] See "Lebanon charges journalists with defamation, fake news," Committee to Protect Journalists, July 12, 2018. Available at: https://cpj.org/2018/07/lebanon-charges-journalists-with-defamation-false-/
[725] "There Is a Price to Pay: The Criminalization of Peaceful Speech in Lebanon," Nov. 15, 2019, Human Rights Watch. Available at: https://www.hrw.org/report/2019/11/15/there-price-pay/criminalization-peaceful-speech-lebanon
[726] "UAE announces $5,500 fine for coronavirus fake news," Al Jazeera, Apr. 18, 2020.
[727] "Cabinet passes resolution on publishing health information about communicable diseases," Emirates News Agency, Apr. 18, 2020.

The history of prosecution under this law indicates that it has been used not only against disinformation, but also misinformation.[728]

g. North America

(i) Canada

In 1992, the Canadian Supreme Court held that a provision criminalizing the spread of false news violated the right to freedom of expression because the restriction was not justified by a pressing and substantial objective as well as being too vague and broad to be proportionate. However, in 2018, Canada passed a law criminalizing disinformation with the intent to influence the result of an election. In 2019, an NGO launched a constitutional challenge of this law. In 2020, at least one governmental official has stated an intent to propose legislation criminalizing disinformation related to COVID-19.

In the case *R v. Zundel (1992)*, Canada's Supreme Court held that a provision criminalizing the spread of false news violated the right to freedom of expression as guaranteed in the Canadian Charter of Rights and Freedoms. The court reasoned that even knowingly false statements are protected by free expression, the provision's purported objective of preserving social harmony was a pretext not rising to the level of a pressing and substantial objective, and the provision was too vague and broad to be proportionate.[729]

Section 181 of Canada's Criminal Code made it a criminal offense for a person to "willfully publish[ ] a statement, tale or news that he knows is false and that causes or is likely to cause injury or mischief to a public interest."[730] The maximum statutory penalty was 2 years in prison. The case originated when Ernst Zundel published a book that denied facts about the Holocaust. He was charged and convicted for spreading false statements under Section 181 of the Canadian Criminal Code. He appealed, asserting a violation of his right to freedom of expression.

[728] George Sadek, *Government Responses to Disinformation on Social Media Platforms: United Arab Emirates.* Global Legal Monitor Entry, Washington, D.C.: Library of Congress, 2020.
[729] *R. v. Zundel*, 2 S.C.R. 731 (Canadian Supreme Court 1992). Available at: https://www.canlii.org/en/ca/scc/doc/1992/1992canlii75/1992canlii75.html
[730] Canadian Criminal Code, R.S.C., 1985, c. C-46, § 181. Available at: https://laws-lois.justice.gc.ca/eng/acts/c-46/FullText.html, archived at https://perma.cc/7EDV-P7WF

The Court followed a multi-step analysis: (1) determining whether the expression conveyed is within the realm of protection of freedom of expression as recognized in the Charter; and then (2) determining if the purpose of the government action or legislation is to restrict expression. If so, the court must determine whether or not that restriction is justified by a pressing and substantial objective and is proportionate.

As regards the first step, the Court asserted that all expression that conveys or attempts to convey a meaning is protected by the provision. Even lies and false statements are protected as expression under the Charter. The court also confirmed that the provision's purpose was to restrict that expression.

Then the Court turned to whether that restriction was justified. It first examined the legislative objective of Section 181 and determined that there was no pressing and substantial objective attached to it: that the preservation of social harmony was just a pretext. Despite finding that Section 181 had no pressing and substantial objective, the Court still undertook an analysis of proportionality. It concluded that Section 181 suffered from vagueness and was overly broad. As such, Section 181 could affect a broad range of expression and speech, and did not pass the proportionality test.

In December of 2018, the Canadian Elections Act was amended to criminalize the spread of false information about candidates, prospective candidates, political party leaders, or public figures with the goal of influencing election results.[731] The prohibition is limited in scope to false statements that someone has committed, been charged with, or being investigated for an offence, or false statements about someone's background. Since the law states that offending action must be undertaken "with the intention of affecting the results of an election," it is likely that the law may only be used to charge the dissemination of disinformation related to electioneering.

In September 2019, the Canadian Constitution Foundation, a libertarian NGO, challenged the constitutionality of the 2018 amendment in court. [732] It argued that the law casted "an

[731] Government of Canada. 2018. "An Act to amend the Canada Elections Act and other Acts and to make certain consequential amendments." Ottawa: Parliament, Dec. 13.
[732] Karanicolas, Michael, "Canada's fake news laws face a Charter challenge. That's a good thing," Ottawa: Ottawa Citizen, Nov. 1, 2019.

exceedingly broad net and could be used to shut down all manner of political speech that is protected by the Charter."[733] No reports have been found of the Canadian courts resolving the challenge.

In 2020, a member of the Prime Minister's Cabinet who serves as President of the Queen's Privy Council stated that, after discussing the issue with leading Members of the Canadian Parliament, he is interested in pursuing legislation that criminalizes "disinformation."[734] He expressed concern for the dissemination of false information relating to the COVID-19 pandemic, which he said Canada's Public Health Agency and Health Canada have been tracking closely.

C. Social Media Company Policies

Leading social media companies have established policies which prohibit users from engaging in certain types of disinformation and, to a lesser extent, misinformation. Violation of one of these policies will generally result in removal of the offending content; repeated violations may result in account termination.

YouTube, Twitter, and Facebook share a large amount of similarity between policies relevant for misinformation and disinformation. These three platforms all ban posts of substantially manipulated media, fake accounts that impersonate others in misleading ways, misinformation or disinformation tending to suppress voting or census participation, and use of multiple accounts in ways that artificially manipulate conversations or mislead users in specified ways. Facebook also generally bans misinformation creating a risk of imminent violence or physical harm. While all three platforms have policies prohibiting impersonation, Twitter notably allows parody or fan accounts that are clearly labeled as such.

The platforms partially diverge when it comes to false information surrounding elections and censuses. Twitter and YouTube both ban posts of false information intended to suppress participation in elections and censuses, and Facebook bans specific categories of election and

[733] "Canadian Constitution Foundation launches constitutional challenge against elections censorship law," Canadian Constitution Foundation, Sept. 18, 2019.
[734] Elizabeth Thompson, "Federal government open to new law to fight pandemic misinformation," CBC News, Apr. 15, 2020.

census misinformation with the same effect, YouTube also bans content that makes false claims about whether candidates or government officials are eligible or ineligible for office without specifying that these claims must be aimed at misleading viewers. Facebook's prohibition includes misrepresentations of the time, locations, qualifications, and effects of voting or census participation, as well as whether or not a candidate is running, without requiring demonstrated intent to engage in voter or census suppression.

The three platforms also have subtly different policies on manipulated media, also known as "deepfakes." Twitter and YouTube both require that the media not only be technologically manipulated in a way that misleads viewers, but also that it do so in a way likely to cause harm (for Twitter, "likely to impact public safety or cause serious harm," and for YouTube: "may pose a serious risk of egregious harm"). Facebook has no parallel requirement of harm, but notably limits its policy to videos created using artificial intelligence or machine learning techniques and specifically exempts "parody or satire" from its ban. In contrast, YouTube and Twitter both give inaccurately translated subtitles as an example of manipulated media eligible for prohibition under their policies, which certainly does not require machine learning to generate.

In response to COVID, YouTube and Facebook essentially reaffirmed the applicability of pre-existing policies to COVID-related misinformation. For Facebook, this included its policy of removing misinformation or "unverifiable rumours" that pose a "risk of imminent violence or physical harm." YouTube references its existing policies on fraud and spam which would ban posts advertising miracle cures. Twitter, in contrast, articulated a list of new guidelines for how to respond to COVID-specific misinformation. These guidelines broadly authorize the removal of false claims of fact about COVID transmission or diagnosis as well as calls for people to disobey health authority recommendations even if made in jest.

1. YouTube

Youtube's policy on "Spam, deceptive practices, and scams," includes prohibitions that relate to disinformation and disinformation: bans on manipulated media, disinformation aimed at voter suppression or census participation suppression, and misinformation on candidate eligibility.

YouTube's policy on "Impersonation" also bans misrepresenting the source of content, via either impersonation of a channel or of another person.[735]

The ban on manipulated media ("deepfakes") has two elements: first, that the content "technically manipulated or doctored in a way that misleads users" beyond merely presenting a video clip "out of context," and second, that it misleads users in a way that "may pose a serious risk of egregious harm." Examples include "[i]naccurately translated video subtitles that inflame geopolitical tensions," videos technologically manipulated "to make it appear that a government official is dead," and "[m]isattributing a 10-year-old video that depicts stuffing of a ballot box to a recent election." [736]

The policy bans disinformation via content "aiming to mislead voters about the time, place, means or eligibility requirements" for either voting or "participating in a census." Listed examples include "[d]eliberately telling viewers an incorrect election date."[737] The policy also bans "[c]ontent that advances false claims related to the technical [i.e. legal] eligibility requirements" of political candidates or governmental officials for serving in office. The wording of this ban, unlike the bans on voter and census suppression attempts, does not require that the content be "aimed at mislead[ing]" viewers. Examples given include claims that a candidate or current government official is ineligible for holding office based on false information about that person's age or citizenship status, or the legal requirements of age or citizenship status.[738]

The policy banning channel impersonation prohibits creators of channels who copying another YouTube channel's "profile, background, or overall look and feel in such a way that makes it look like someone else's channel," even if the end result is not entirely identical to the impersonated channel. Examples include using the same name as another channel but inserting an extra space into the name.[739]

[735] "Rules and Policies: Community Guidelines," Youtube, as of Aug 11, 2020. Available at: https://www.youtube.com/howyoutubeworks/policies/community-guidelines/#community-guidelines
[736] *Id.*
[737] *Id.*
[738] *Id.*
[739] *Id.*

Similarly, the ban on personal impersonation applies to content "intended to look like someone else is posting it." Examples include setting up a channel with a person's name and image and then pretending to be that person while posting content to that channel or posting comments on other channels.[740]

YouTube has no particular policy dedicated to misinformation around COVID-19, but has noted that it will continue to remove flagged videos that violate pre-existing policies, including by "discourag[ing] people from seeking medical treatment or claim[ing] harmful substances have health benefits."[741] YouTube's policy on scams does ban content that makes "exaggerated promises, such as . . . that a miracle treatment can cure chronic illnesses." However, there is no pre-existing policy that would clearly ban content discouraging medical treatment.[742]

2. Twitter

Twitter has policies banning synthetic and manipulated media likely to cause harm, disinformation intended to suppress participation in a civic process like elections, and impersonation of others in ways intended to or actually causing confusion, as well as a broader prohibition on using "coordinated activity" to "artificially influence conversations."[743]

Twitter's policy on manipulated media prohibits users from "deceptively shar[ing] synthetic or manipulated media that are likely to cause harm." The policy lists three elements: if the content is (1) "significantly and deceptively altered or fabricated," (including by overdubbing audio or modifying subtitles, or otherwise doctoring the content "to change its meaning"); (2) "shared in a deceptive manner," (including whether the context suggests " a deliberate intent to deceive people about the nature of the origin of the content," such as by "claiming that it depicts reality) and (3) "likely to impact public safety or cause serious harm." This last criterion of

740 *Id.*
741 "Coronavirus disease 2019 (COVID-19) updates," YouTube Help Center, last updated July 13, 2020. Available at: https://support.google.com/youtube/answer/9777243?p=covid19_updates&visit_id=637327908310564492-3826126572&rd=1
742 "Rules and Policies: Community Guidelines," YouTube, as of Aug. 11, 2020. Available at: https://www.youtube.com/howyoutubeworks/policies/community-guidelines/#community-guidelines
743 "Twitter Rules and policies," Twitter Help Center, as of Aug. 11, 2020. Available at: https://help.twitter.com/en/rules-and-policies/twitter-rules

likelihood of serious harm includes threats to someone's physical safety, risks of "mass violence or widespread civil unrest," or threats to someone's ability to engage in free expression or participate in civic events, such as "voter suppression or intimidation." Immediacy of likely resulting harm makes removal more likely. If all three elements are met, then the content is "likely to be removed." If only the first and third are met (significant deceptive alteration and likelihood of harm), the content "may be removed," but may instead merely be labeled as manipulated media.[744]

Twitter's "Civic Integrity" policy prohibits using Twitter "for the purpose of manipulating or interfering in elections or other civic processes," including by posting "content that may suppress participation or mislead people about when, where, or how to participate." Covered civic processes include political elections, censuses, and "[m]ajor referenda and ballot initiatives." This policy prohibits (1) "false or misleading information about how to participate," including requirements for participation, (2) "false or misleading information intended to intimidate or dissuade" participation, including misleading claims about long lines, and (3) fake accounts which "misrepresent their affiliation [ ] to a candidate, elected official, political party . . . or government entity." Therefore, this policy appears to prohibit both misinformation and disinformation with a tendency to suppress civic participation. However, the policy specifically excludes "inaccurate statements about an elected or appointed official, candidate, or political party" from being banned under this policy.[745]

Twitter's "Impersonation" policy bans users from "impersonat[ing] individuals, groups, or organizations in a manner that is intended to or does mislead, confuse, or deceive others."[746] However, Twitter affirmatively allows users to operate "parody . . . commentary, or fan accounts,"

[744] "Synthetic and maniupalted media policy," Twitter Help Center, as of Aug 11, 2020. Available at: https://help.twitter.com/en/rules-and-policies/manipulated-media
[745] "Civic integrity policy," Twitter Help Center, May 2020. Available at: https://help.twitter.com/en/rules-and-policies/twitter-impersonation-policy
[746] "Impersonation policy," Twitter Help Center, as of Aug 11, 2020. Available at: https://help.twitter.com/en/rules-and-policies/twitter-impersonation-policy

as long as both the account name and account "bio" clearly indicate that the account is a parody, commentary, fake, or fan account.[747]

Twitter's "Platform manipulation and spam" policy more broadly prohibits the use of Twitter "in a manner intended to artificially amplify or suppress information or engage in behavior that manipulates or disrupts people's experience on Twitter." This includes misleading others by "operating fake accounts" and using multiple accounts—or coordinating with others —to "amplify or disrupt conversations," as well as using spamming or other techniques to artificially inflate follower counts or hijacking hashtags in order to "drive traffic or attention" to other accounts, websites, or initiatives. However, the policy clarifies that operating an account under a pseudonym or as a "parody, commentary, or fan account" does not violate the rules.[748]

Twitter's response to COVID included broadening its definition of harm and articulating COVID-19-specific guidance for misinformation. Twitter announced that it would continue to prioritize the removal of content containing "a clear call to action that could directly pose a risk to people's health or well-being," but emphasized that it would not be removing every Tweet with "incomplete or disputed information about COVID-19." However, Twitter said it would remove (1) attempts to convince people not follow health authority recommendations, including tweets like "social distancing is not effective;" (2) descriptions of alleged cures for COVID-19 "known to be ineffective . . . even if made in jest;" (3) "[d]enial of established scientific facts about transmission" of COVID-19 or "false or misleading information" about how to diagnose COVID.[749]

3. Facebook

Facebook's Community Standards prohibit misrepresentation of one's identity, the posting of manipulated videos created using machine learning or artificial intelligence. Facebook also

[747] "Parody, newsfeed, commentary, and fan account policy", Twitter Help Center, as of Aug. 11, 2020. Available at: https://help.twitter.com/en/rules-and-policies/parody-account-policy
[748] "Platform manipulation and spam policy," Twitter Help Center, Sept. 2019. Available at: https://help.twitter.com/en/rules-and-policies/platform-manipulation
[749] "An update on our continuity strategy during COVID-19," Twitter, March 16, 2020, last updated Apr. 1, 2020. Available at: https://blog.twitter.com/en_us/topics/company/2020/An-update-on-our-continuity-strategy-during-COVID-19.html

prohibits users from misleading people about the origin or source of content, including using multiple fake accounts to do so or doing so on behalf of a foreign government. [750] Finally, Facebook prohibits generalized "misinformation and unverifiable rumors" only if they "contribute to the risk of imminent violence or physical harm."

Facebook's "Manipulated media" policy prohibits the posting of videos that (1) have been altered in ways "not apparent to an average person" that "would likely mislead an average person to believe that a subject of the video said words that they did not say, and (2) have been created using artificial intelligence or machine learning. The policy specifies that it does not ban "parody or satire," or other content edited to omit or change the order of words.

Facebook's "Misrepresentation" policy prohibits users from misrepresenting their identities, which includes providing a false name or date of birth as well as impersonating others. The policy includes an unexplained statement that impersonation includes the posting of "imagery that is likely to deceive the public as to the content's origin" as long as "[t]he entity or an authorised representative objects to the content, and [c]an establish a risk of harm to members of the public." However, it is unclear what is meant by "the entity," so the practical effect of the statement might not be as broad as it first appears.

Facebook's "Inauthentic Behavior" policy prohibits "inauthentic behavior," defined as misleading people on Facebook about "the identity, purpose or origin of the entity they represent" or about a piece of content's source or origin. Users also may not engage in this kind of behavior in a coordinated manner using fake accounts or engage in inauthentic behavior "conducted on behalf of a foreign or government actor."

Under Facebook's "Coordinating harm and publicizing crime" policy, posts that suppress voting or census participation by misrepresent election or information in specific ways (mispresenting the time, place, method, qualifications and effects of or for participation or of whether or not a candidate is running) are banned and may be removed rather than merely de-ranked under Facebook's general "false news" policy.

[750] Community Standards, Facebook, as of Aug 11, 2011. Available at: https://www.facebook.com/communitystandards/

For other forms of misinformation, under the heading of "false news," Facebook has generally merely reduces distribution of such content rather than removing it. However, Facebook's policy on bans and provides for the removal of "[m]isinformation and unverifiable rumours that contribute to the risk of imminent violence or physical harm."

Facebook has not adopted a specific policy in response to COVID-19, but has emphasized that it is enforcing its relevant existing policies "prohibiting the coordination of harm, the sale of medical masks and related goods . . . and misinformation that contributes to the risk of imminent violence or physical harm." Under Facebook's policies on "Violence and criminal behavior," the platform prohibits "content aimed at deliberately deceiving people to gain an unfair advantage or deprive another of money, property or legal right."

D. Conclusions

International human rights jurisprudence evinces a reluctance to ratify laws that criminalize the spread of false information as compatible with the right to freedom of expression. After analyzing the comments by non-binding authorities on freedom of expression in international human rights, the main source of that reluctance appears to be how misinformation offences could be easily harnessed by state powers to stifle dissent. In that line, international human rights organizations constantly warn about false-information crimes being used to persecute political opposition. Perpetrators of these crimes are often journalists who publish stories critical of the government.

When international courts rule on criminal laws punishing the publication of false information, they have tended to find their applications to violate free expression. When these laws sound in defamation, the international jurisprudential principle that defamation may be civilly but not criminally punished stands in tension with the large number of states that have criminalized information.

However, the three international court cases analyzing the freedom of expression implications of more generalized criminal disinformation or misinformation laws have not expressed in their holdings that criminally punishing the spread of false information can never

comply with international law. Instead, these courts took issue with how particular laws were drafted or applied, finding that the vague provisions or domestic court's placing the burden of proving truth or falsity on the defendant violated free expression, chilling free speech and public debate.

In two instances, the laws in question were phrased too vaguely, without *mens rea* requirements for falsity. Thus, the laws violated the first prong of the tri-partite test: that the restriction on free expression be prescribed by law with sufficient particularity to give notice of what will be penalized. These two laws either simply criminalized spreading false statements, or did so when they could cause fear to the population or threaten abstract concepts, without specifying mens rea requirements as to the falsity of the information, and thus were found to fall short of the standards of freedom of expression contained in international law.

Domestic courts have followed similar criteria, with varying degrees of leniency towards the vagueness of the law's language. While international tribunals and some national courts have found that broad prohibitions have an unacceptable chilling effect, especially when applied to political speech, some national courts have upheld even broadly phrased offences as constitutional with respect to national rights to free expression.

In general, African nations seem to have taken the most action with regard to not only passing criminal laws aimed at reducing the spread of false information, but also in exercising judicial oversight over the constitutionality of these laws. The regions of Middle East and Asia have taken similarly aggressive actions toward criminalizing the spread of false information, but with fewer instances of judicial oversight. In comparison, countries in Central and South America and Oceania have taken fewer and generally less aggressive actions on misinformation or disinformation.

In Europe and the Commonwealth of Independent States, there is a mixture of state actions. When EU members take action against misinformation, they tend to do it through non-criminal means, such as intermediary liability (with Germany's NetzDG leading the way), awareness campaigns, and task forces.

While definitions of false information vary among states and regions. there is still a discernible trend in terms of their focus. Most pieces of legislation that penalize the spread of false

information criminalize the spread of disinformation, while leaving open the question of whether misinformation may also be prosecuted. In various instances, laws phrased in ways that seem to target disinformation has resulted in charges against social media users that seem to have inadvertently shared false information.

Additionally, most laws criminalizing the spread of false information require that the false communication either cause or be likely to cause harm to public interests, whether that be public disorder, an undermining of people's trust in government, or a threat to national security and public health. Ultimately, in many jurisdictions the laws are left vague and it is incumbent on courts to determine whether a specific dissemination of false information presents a sufficient harm to public interests to justify prosecution, or how intention to deceive, or lack thereof, should play a role in convictions and sentencing.

F. Appendix – Data Disaggregated by Region

In the below tables, percentages highlighted in yellow fall outside of one standard deviation from the mean for that metric (and are thus statistically atypical), whereas percentages highlighted in green are within one standard deviation from the mean (and are thus statistically typical).

Overall: Whether and What Kind of Action Was Considered or Taken										
Region	**Total**	**Action**	**Action**	**No action**	**No action**	**Unsure**	**Of Action: Criminal**	**Of Action: Criminal**	**Of Action: Civil**	**Of Action: Civil**
Africa	54	27	50%	25	46%	2	25	93%	2	7%
Asia	24	19	79%	4	17%	1	14	74%	5	26%
Caribbean	13	4	31%	9	69%	0	4	100%	0	0%
Central & South America	19	6	32%	13	68%	0	2	33%	4	67%
CIS	10	7	70%	3	30%	0	5	71%	2	29%
Europe (non-EU)	15	8	53%	7	47%	0	5	63%	3	38%
Europe (EU)	27	15	56%	12	44%	0	6	40%	9	60%
Middle East	15	10	67%	5	33%	0	10	100%	0	0%
North America	3	3	100%	0	0%	0	1	33%	2	67%
Oceania	14	2	14%	12	86%	0	0	0%	2	100%

Total	194	101	52%	90	46%	3	72	71%	29	29%
Standard Deviation			25%		26%			33%		33%

COVID-19: Whether and What Kind of Action Was Considered or Taken						
Region	**Of Action: COVID**	**Of Action: COVID**	**Of COVID Action: Criminal**	**Of COVID Action: Criminal**	**Of COVID Action: Civil**	**Of COVID Action: Civil**
Africa	8	30%	7	88%	1	13%
Asia	4	21%	4	100%	0	0%
Caribbean	2	50%	2	100%	0	0%
Central & South America	1	17%	1	100%	0	0%
CIS	3	43%	2	67%	1	33%
Europe (non-EU)	1	13%	1	100%	0	0%
Europe (EU)	3	20%	1	33%	2	67%
Middle East	2	20%	2	100%	0	0%
North America	1	33%	1	100%	0	0%
Oceania	1	50%	0	0%	1	100%
Total	24	24%	21	88%	5	21%
Standard Deviation		14%		35%		35%

Detail: What Type of Criminal or Civil Action Was Considered or Taken								
Region	**Of Criminal Action: Defamation**	**Of Criminal Action: Defamation**	**Of Civil Action: Individual Liability**	**Of Civil Action: Individual Liability**	**Of Civil Action: Corporate liability**	**Of Civil Action: Corporate Liability**	**Of Civil Action: Campaign /Task Force**	**Of Civil Action: Campaign /Task Force**
Africa	2	25%	0	0%	0	0%	2	100%
Asia	0	0%	0	0%	4	80%	1	20%
Caribbean	2	100%	0	0%	0	0%	0	0%
Central & South America	0	0%	1	25%	1	25%	2	50%
CIS	1	33%	1	50%	1	50%	0	0%
Europe (non-EU)	2	200%	1	33%	1	33%	1	33%
Europe (EU)	0	0%	0	0%	3	33%	6	67%
Middle East	1	50%	0	0%	0	0%	0	0%
North America	0	0%	0	0%	0	0%	2	100%
Oceania	0	0%	0	0%	1	50%	1	50%
Total	8	33%	3	10%	11	38%	15	52%

Standard Deviation		65%		18%		28%		38%

Below is a list of relevant metrics and the ranges of typicality pulled from the tables above:

Total

- 52% have taken or considered action against misinformation.
    - Typical in range of 27-77%
- 71% of that action includes criminalization of the spread of misinformation, false news, or other synonymous activities.
    - Typical in range of 38-100%
- 24% of all the action taken was in response to the COVID-19 pandemic.
    - Typical in range of 10-38%
- 88% of action in response to the COVID-19 pandemic includes criminalization.
    - Typical in range of 53-100%
- 11% of all actions taken or considered to criminalize misinformation is done through the use of anti-defamation laws.
    - Typical in range of 0-29%

Africa

- 50% have taken or considered action against misinformation.
    - Typical
- 93% of that action includes criminalization of the spread of misinformation.
    - Typical
- 30% of all the action taken was in response to the COVID-19 pandemic.
    - Typical
- 88% of action in response to the COVID-19 pandemic includes criminalization.
    - Typical
- 8% of all actions taken or considered to criminalize misinformation is done through the use of anti-defamation laws.
    - Typical

Asia

- 79% have taken or considered action against misinformation.
    - Not typical, high
- 74% of that action includes criminalization of the spread of misinformation.
    - Typical
- 21% of all the action taken was in response to the COVID-19 pandemic.
    - Typical
- 100% of action in response to the COVID-19 pandemic includes criminalization.
    - Typical
- 0% of all actions taken or considered to criminalize misinformation is done through the use of anti-defamation laws.
    - Typical

Caribbean

- 31% have taken or considered action against misinformation.
    - Typical
- 100% of that action includes criminalization of the spread of misinformation.
    - Typical
- 50% of all the action taken was in response to the COVID-19 pandemic.
    - Not typical, high
- 100% of action in response to the COVID-19 pandemic includes criminalization.
    - Typical
- 50% of all actions taken or considered to criminalize misinformation is done through the use of anti-defamation laws.
    - Not typical, high

Central and South America

- 32% have taken or considered action against misinformation.
    - Typical
- 33% of that action includes criminalization of the spread of misinformation.
    - Not typical, low

- 17% of all the action taken was in response to the COVID-19 pandemic.
    - Typical
- 100% of action in response to the COVID-19 pandemic includes criminalization.
    - Typical
- 0% of all actions taken or considered to criminalize misinformation is done through the use of anti-defamation laws.
    - Typical

CIS

- 70% have taken or considered action against misinformation.
    - Typical
- 71% of that action includes criminalization of the spread of misinformation.
    - Typical
- 43% of all the action taken was in response to the COVID-19 pandemic.
    - Not typical, high
- 67% of action in response to the COVID-19 pandemic includes criminalization.
    - Typical
- 20% of all actions taken or considered to criminalize misinformation is done through the use of anti-defamation laws.
    - Typical

Europe (Non-EU)

- 53% have taken or considered action against misinformation.
    - Typical
- 63% of that action includes criminalization of the spread of misinformation.
    - Typical
- 13% of all the action taken was in response to the COVID-19 pandemic.
    - Typical
- 100% of action in response to the COVID-19 pandemic includes criminalization.
    - Typical

- 40% of all actions taken or considered to criminalize misinformation is done through the use of anti-defamation laws.
    - Not typical, high

Europe (EU)

- 56% have taken or considered action against misinformation.
    - Typical
- 40% of that action includes criminalization of the spread of misinformation.
    - Typical
- 20% of all the action taken was in response to the COVID-19 pandemic.
    - Typical
- 33% of action in response to the COVID-19 pandemic includes criminalization.
    - Not typical, low
- 0% of all actions taken or considered to criminalize misinformation is done through the use of anti-defamation laws.
    - Typical

Middle East

- 67% have taken or considered action against misinformation.
    - Typical
- 100% of that action includes criminalization of the spread of misinformation.
    - Typical
- 20% of all the action taken was in response to the COVID-19 pandemic.
    - Typical
- 100% of action in response to the COVID-19 pandemic includes criminalization.
    - Typical
- 10% of all actions taken or considered to criminalize misinformation is done through the use of anti-defamation laws.
    - Typical

North America

- 100% have taken or considered action against misinformation.

    - Not typical, high
- 33% of that action includes criminalization of the spread of misinformation.
    - Not typical, low
- 33% of all the action taken was in response to the COVID-19 pandemic.
    - Typical
- 100% of action in response to the COVID-19 pandemic includes criminalization.
    - Typical
- 0% of all actions taken or considered to criminalize misinformation is done through the use of anti-defamation laws.
    - Typical

Oceania

- 14% have taken or considered action against misinformation.
    - Not typical, low
- 0% of that action includes criminalization of the spread of misinformation.
    - Not typical, low
- 50% of all the action taken was in response to the COVID-19 pandemic.
    - Not typical, high
- 0% of action in response to the COVID-19 pandemic includes criminalization.
    - Not typical, low
- 0% of all actions taken or considered to criminalize misinformation is done through the use of anti-defamation laws.
    - Typical

VII. DEFAMATION ONLINE

Modern defamation law is situated within a conflict embedded into international human right law (IHRL), between the right to freedom of expression and the right to reputation. Both are core rights protected by IHRL. The question of how to balance them has been the source of significant disagreements under domestic and human rights law.

Broadly speaking, defamation refers to speech that infringes upon another's reputation. The breadth of this definition lends itself to multiple interpretations. For instance, in some jurisdictions, a true statement cannot be defamatory, while in others, even truth does not operate as a complete defense.

The difficulty of balancing the rights to free speech and reputation predate the controversies around defamation on the internet. In any medium, an absolute right to freedom of expression would militate against any restriction on its exercise, while and an absolute focus on the right to reputation or privacy would severely circumscribe the boundaries of free speech.

The internet only heightens these tensions. In a world of instantaneous mass dissemination, one might argue for more stringent controls on defamatory speech. However, it can also be argued that the internet offers the possibility of near-simultaneous responses to allegedly defamatory material and thus, speech on the internet should be *less* sanctioned than speech on more traditional forms of media.[751] The tension between some national practice and international human rights norms is particularly pronounced in cases of religious defamation and the defamation of heads of state.

The following sections analyze how restrictions on defamatory speech alternately comply and violate the international right to freedom of expression in four ways.

First, this chapter briefly indicates which treaties enshrine the rights to reputation and privacy as principles of international human rights law.

[751] *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, Frank La Rue, U.N. Human Rights Council, A/HRC/17/27, May 16, 2011, at Para 27. Available at: https://www2.ohchr.org/english/bodies/hrcouncil/docs/17session/A.HRC.17.27_en.pdf

Second, this chapter analyzes trends in international legal jurisprudence on defamation. These include: (i) that to be defamatory, statements must be factual rather than opinion-based; (ii) that despite state practice to the contrary, international courts consider criminalization of defamation to be too severe a form of restriction to comply with free expression, (iii) that free expression requires governments to show particular restraint in restricting criticism of public figures and heads of state, and (iv) that disagreement exists over the permissibility of criminal blasphemy laws.

Third, this chapter surveys national laws and identify regional legislative trends in Europe, the Americas, the Middle East, Africa, and the Asia Pacific. In each region, we trace how state practice reflects and informs the abovementioned issues, including laws imposing heightened penalties on defamation criticizing public officials or heads of state.

Fourth, this chapter summarizes the approaches taken to defamation by social media platforms. Most major platforms do not mention defamation in their community guidelines; instead, they process legal take-down requests on the basis of local illegality. For instance, if a defamed person submits a court order declaring content illegal in a particular country, platforms will generally block access to that content in that country.

A. Relevant International Human Right Law (IHRL) Treaties

The right to freedom of expression is integral to IHRL.[752] As discussed in greater detail in the chapter on IHRL architecture, this right is firmly established in many treaties around the world. However, international law also protects a right to reputation and a right to privacy.

Global treaties and IHRL instruments clearly enshrine human rights to reputation and privacy. The Universal Declaration of Human Rights asserts that "no one shall be subjected to […]

[752] For instance, the Human Rights Committee has described the freedoms of expression and opinion as being "indispensable conditions for the full development of the person." *General comment no. 34, Article 19, Freedoms of opinion and expression*, UN Human Rights Committee (HRC), Sept. 12, 2011.

attacks upon his honour and reputation."[753] Article 17 of the ICCPR[754] protects individuals against "unlawful attacks" on their "honour and reputation." Article 19 of the ICCPR allows for restrictions on the freedom of expression in order to ensure "respect of the rights or reputations of others."[755]

Regional human rights treaties echo the importance of rights of reputation and privacy. Article 8 of the European Convention on Human Rights (ECHR) protects a right to private and family life, while Article 10 of the ECHR notes that the freedom of expression may be limited "for the protection of the reputation or rights of others."[756] Article 11 of the American Convention on Human Rights (ACHR) recognizes a right to privacy and reputation.[757] In contrast, the Arab Charter on Human Rights (AC) does not specifically reference defamation.[758] The African Charter on Human and Peoples Rights (AfCHR) also does not specifically mention privacy or reputation, but it does emphasize that the "rights and freedoms of each individual shall be exercised with due regard to the rights of others."[759]

B. International Legal Jurisprudence

In this section, we analyze opinions from international courts on the following issues. First, whether defamatory statements must be factual or if they can be opinion-based. Second, despite state practice to the contrary, international courts consider criminalization of defamation to be too severe a form of restriction to comply with free expression. Third, that free expression requires

[753] UDHR, Article 12.
[754] ICCPR, Article 17: "1. No one shall be subjected to arbitrary or unlawful interference with his privacy, family, home or correspondence, nor to unlawful attacks on his honour and reputation. 2. Everyone has the right to the protection of the law against such interference or attacks."
[755] ICCPR, Article 19(3)(a): "3. The exercise of the rights provided for in paragraph 2 of this article carries with it special duties and responsibilities. It may therefore be subject to certain restrictions, but these shall only be such as are provided by law and are necessary: (a) For respect of the rights or reputations of others . . . ."
[756] ECHR, Article 10.
[757] ACHR, Article 11.
[758] Arab Charter on Human Rights, Article 26. Available at: http://hrlibrary.umn.edu/instree/loas2005.html
[759] AfCHR, Article 27(2). Available at: https://www.achpr.org/legalinstruments/detail?id=49

governments to show particular restraint in restricting criticism of public figures and heads of state. Fourth, we explore the relationship between religious blasphemy laws and defamation.

1. Fact vs. Opinion

One of the important points of controversy in defamation concerns what types of speech may be considered defamatory. The degree to which opinion is considered protected speech has been extensively litigated upon. However, human rights bodies have generally agreed that defamation should involve an untrue statement. For instance, the Human Rights Committee's General Comment No. 34 interpreting Article 19 of the ICCPR notes that opinion should not be within the ambit of defamation since opinions are, by definition, impossible to verify.[760]

Since opinions are assertions whose truth cannot be verified, a number of courts have ruled that criminalizing a statement of opinion as defamation violates freedom of expression. The ECtHR established this doctrine in *Lingens v. Austria,* emphasizing the need for "a careful distinction . . . between facts and value judgements," as facts can be proved but value judgements cannot.[761] The court held a criminal penalty (a fine) unnecessary and disproportionate in part because the domestic court had required the defendant to prove the truth of his statement, which was impossible because his statement was a value-judgment, i.e. an opinion. The ECtHR declared that this requirement itself violated freedom of expression. The ECtHR has applied similar reasoning to insults, such as a journalist calling a politician an "idiot" in *Oberschlick v. Austria*, despite noting that the journalist explained an accurate factual basis underlying this insult.[762]

Three national courts in South America have come to similar conclusions suggesting that in order to be criminalized, defamatory statements must be statements of fact rather than opinion. In 2016, the Superior Court of Lima in Peru ruled that a journalist could not be charged with defamation for an opinion column.[763] In 2014, the Federal Supreme Court of Brazil ruled that the

760 *General comment no. 34,* at Para. 47.
761 *Lingens v. Austria*, Application No. 9815/82 (ECtHR 1986).3
762 *Oberschlick v. Austria (no. 2)*, App. No. 47/1996/666/852 (ECtHR 1997).
763 Judgement, No. 14156-2014, (Corte Superior de Justicia de Lima, Aug. 29, 2016. Available at: http://legis.pe/wp-content/uploads/2016/09/Lee-aqu%C3%AD-la-sentencia-de-segunda-instancia-queabsuelve-a-Rafo-Le%C3%B3n-Legis.pe_.pdf

right to "have and share opinions" was integral to the right of freedom of expression and noted that unless such a right were protected the freedom of the press would be sharply attenuated.[764] In 2013, the Supreme Court of Colombia held that a journalist could not have a criminal case brought against him based on his use of "disrespectful" language about a former governor because these comments were based on the columnist's perceptions and did not contain factual elements.[765]

2. Criminalizing Defamation

International human rights law has tended to look askance on the phenomenon of *criminalizing* defamation, whether online or offline. However, there is substantial divergence between international jurisprudence and state practice on this point.

An association of international human rights law experts named criminal defamation laws as one of the "ten key threats to freedom of expression" in 2010.[766] There have been numerous calls to decriminalize defamation.[767] For instance, the Human Rights Committee has called for decriminalizing defamation, asserting that *"*imprisonment is never an appropriate penalty."[768]

European human rights bodies and courts have consistently found that criminalization of defamation is a disproportionate restriction that violates free expression. The European Court of Human Rights (ECtHR) has not officially denounced or banned criminal defamation laws *per se*, but has criticized the excessive use of criminal defamation laws, and repeatedly found their use to be a disproportionate restriction on free expression, in violation of Article 10 of the European Convention on Human Rights (ECHR). The ECtHR accepts criminal penalties rather than civil

764 Judgment of Dec. 17, 2014, 16.329 MC/CE (Supremo Tribunal Federal do Brasil), published on Feb. 2, 2015.
765 *Gonzalez v. Serrano*, Rad. No. 38.909 (Corte Suprema de Justicia de Colombia, July 10, 2013).
766 *Tenth Anniversary Joint Declaration: Ten Key Challenges to Freedom of Expression in the Next Decade*, U.N. Special Rapporteur on Freedom of Opinion and Expression, the OSCE Representative on Freedom of the Media, the OAS Special Rapporteur on Freedom of Expression, & the ACHPR Special Rapporteur on Freedom of Expression and Access to Information, Feb. 3, 2010. Available at: http://www.oas.org/en/iachr/expression/showarticle.asp?artID=784&lID=1
767 *Concluding observations on Italy*, U.N. Human Rights Committee, CCPR/C/ITA/CO/5, May 1, 2017; *Concluding observations on the Former Yugoslav Republic of Macedonia*, U.N. Human Rights Committee, CCPR/C/MKD/CO/2, Apr. 17, 2008.
768 *General comment no. 34,* U.N. Human Rights Committee, at Para. 47.

restrictions only in extraordinary circumstances in which the speech at issue impairs other fundamental rights, such as "in the case of hate speech or incitement to violence."[769]

Similarly, the Inter-American system has repeatedly denounced the criminalization of defamation. In 2000, the Inter-American Commission adopted the Declaration of Principles on Freedom of Expression, which provided that defamation laws protecting reputations of public figures should only impose civil sanctions and noted that "*desacato* laws," which penalize offensive expressions directed at public officials, restrict freedom of expression.[770] The Inter-American Court of Human Rights (IACtHR) has also taken a strong stance against criminal defamation, especially in the context of journalists, and consistently adopts "precautionary measures," in which the court calls on member states to take immediate corrective action, to aid journalists facing imprisonment.[771] In 2012, this measure proved successful in ensuring that Ecuadorian journalists for the *El Universo* newspaper were not imprisoned for making allegedly defamatory statements against the Ecuadorian President.[772]

The African Commission has also called for the decriminalization of defamation. The African Court of Human and Peoples' Rights ruled in *Konaté v. Burkina Faso* in 2014 that criminal charges for libel, unless exceptional circumstances are present, are disproportionate responses that therefore violate freedom of expression.[773] The court defined those exceptional circumstances as speech that constitutes incitement to violence or hate speech.[774]

The 1996 United Nations Educational Scientific and Cultural Organization Declaration of Sana'a, mainly drafted by Arab delegates, states that "disputes involving media and/or the media professionals in the exercise of their profession … should be tried under civil and not criminal

769 Merita Kettunen, *Legitimizing European Criminal Law: Justification and Restrictions*, Springer Nature, Nov. 8, 2019.
770 *Declaration of Principles on Freedom of Expression*, Inter-American Commission on Human Rights, 2000. Available at: http://www.oas.org/en/iachr/expression/showarticle.asp?artID=26
771 Alexandra Ellerbeck, "Inter-American Human Rights System, campaigns against defamation laws keep journalists from jail in Americas," Committee to Project Journalists, Dec. 15, 2015.
772 La CIDH otorga medidas cautelares a directivos de El Universo y Emilio Palacio, El Universo, Feb. 21, 2012. Available at: https://www.eluniverso.com/2012/02/21/1/1355/cidh-otorga-medidas-cautelares-directivos-universo-emilio-palacio-II.html
773 *Konaté v. Burkina Faso*, App. No. 004/2013 (Afr. Court Hum. Peoples Rights, Dec. 5, 2014). Available at: http://www.ijrcenter.org/wp-content/uploads/2015/02/Konate-Decision-English.pdf
774 *Id.*

codes and procedures."[775] Despite this standard, a study commissioned by the OSCE Representative on Freedom of the Media found that "42 of the 57 OSCE participating States maintain general criminal defamation laws," and that "[i]n the vast majority of these cases, defamation and or insult carries a potential penalty of imprisonment."[776]

On the other hand, state practice may indicate an emergent trend towards decriminalizing defamation. Since 2000, over thirty states have taken steps towards doing away with criminalizing defamation and over ten countries have abolished the imposition of prison sentences for defamation.[777]

3. Criticizing Public Figures

The degree to which public officials may be protected against allegedly defamatory statements has been the subject of significant disagreement between international human rights law and domestic laws. Many international and national courts have emphasized that states must refrain from prohibiting criticism of public officials. Many states follow this principle. However, a number of states prescribe more severe penalties if the person being defamed is an official figure, such as via *lese majeste,* heads of state, and *desacato* laws.

In general, human rights law has called for public officials to be subject to a heightened standard for defamation: that is, that they should be willing to tolerate sharper criticism than private persons do. In general, human rights authorities appear to be motivated by the dangers that accompany the possibility of the "chilling effect" on speech when the state's power is mobilized against an individual on the basis that a public official has been criticized.[778]

Both regional and supra-regional organizations have taken this position. For example, the Human Rights Committee has instructed that "the mere fact that forms of expression are

[775] Declaration of Sana'a, United Nations Educational Scientific and Cultural Organization, 1996, at 21.
[776] Griffen et al,"Defamation and Insult Laws in the OSCE Region: A Comparative Study," Organization for Security and Co-operation in Europe, International Press Institute, March 2017. Available at: https://ipi.media/criminal-defamation-unduly-limits-media-freedom/
[777] Amicus Curiae Brief of Intl. Human Rights Clinic at Yale Law School, *In re Emilio Palacio Urrutia et al*.
[778] Jane Kirtley, "Criminal Defamation: An 'Instrument of Destruction,'" *Ending the Chilling Effect: Working to Repeal Criminal Libel and Insult Laws*, eds. Ana Karl Reiter and Hanna Vuokko, 2004. Available at: http://www.osce.org/fom/13573

considered to be insulting to a public figure is not sufficient to justify the imposition of penalties"[779] and that, "with regard to comments about public figures," states should avoid penalizing "untrue statements that have been published in error but without malice."[780] Moreover, the Human Rights Committee has also stated that "a public interest in the subject matter of the criticism should be recognised as a defence." [781]

Similarly, in the *Nyanzi* case discussed below, the United Nations Human Rights Council Working Group on Arbitrary Detention (WGAD) relied on the General Comment 34 to hold that "the right to freedom of expression in article 19 (2) of the Covenant includes the right of individuals to criticize or openly and publicly evaluate their Governments without fear of interference or punishment."[782] The Organization of American States has reiterated the position that public figures should be willing to accept heightened scrutiny. [783] The UN Special Rapporteur has stated that reporting on the government should be unrestricted.[784] The OSCE Representative on Freedom of the Media, together with similar figures in the UN and Inter-American systems, declares that "defamation laws should reflect . . . the principle that public figures are required to accept a greater degree of criticism than private citizens," and that states should repeal "laws which provide special protection for public figures, such as *desacato* laws."[785] *Desacato* laws prohibit insulting, threatening, or injuring a public functionary.[786]

[779] *General comment no. 34,* at Para 38.
[780] Concluding observations on the United Kingdom of Great Britain and Northern Ireland, U.N. Human Rights Committee, CCPR/C/GBR/CO/6, July 21, 2008.
[781] *General comment no. 34,* at Para 47.
[782] *Opinion No. 57/2017 concerning Stella Nyanzi*, A/HRC/WGAD/2017/57, at Para 53 (Human Rights Council Working Group on Arbitrary Detention Aug. 2017).
[783] *Annual Report of the Special Rapporteur for Freedom of Expression*, Office of the Special Rapporteur for Freedom of Expression Inter-American Commission on Human Rights, 2004, Chapter 7. Available at: http://www.cidh.oas.org/relatoria/showarticle.asp?artID=459&lID=1
[784] *Promotion and protection of the right to freedom of opinion and expression: Note by the Secretary-General*, U.N. General Assembly, A/66/290, Aug. 10, 2011, at Para 42. Available at: www.un.org/Docs/journal/asp/ws.asp?m=A/66/290
[785] *Joint Declaration about Censorship by Killing and Defamation*, The UN Special Rapporteur on Freedom of Opinion and Expression, the OSCE Representative on Freedom of the Media and the OAS Special Rapporteur on Freedom of Expression, 2000. The Special Rapporteur on Freedom of Expression and Access to Information of the African Commission on Human and People's Rights has joined more recent, similar statements.
[786] *Chapter IV - Laws on Contempt, Compulsory Membership, and Murder of Journalists*, Special Rapporteurship For Freedom Of Expression, Organization of American States, 2020. Available at: http://www.oas.org/en/iachr/expression/showarticle.asp?artid=633&lid=1

This approach of giving greater latitude to criticism of public figures has been reiterated by international human rights courts. For instance, in *Lingens* v. *Austria,* the ECtHR held that "[t]he limits of acceptable criticism are . . . wider as regards a politician as such than as regards a private individual;" since a politician knowingly opens up "his every word and deed" to close public scrutiny, he must show more tolerance towards criticism.[787] The ECtHR has also applied this towards inflammatory insults. In the case of *Eon v. France*, an individual ran afoul of "insult" laws for calling the French President Sarkozy a "sad prick;" the ECtHR held that his criminal penalty—despite only being a suspended fine of 30 euros—was disproportionate and therefore unnecessary in a democratic society.[788] In a similar vein, the IACtHR has also ruled that public officials should be willing to undergo greater scrutiny than private individuals.[789]

Some national courts have also adopted this position, including courts in Latin America and South America. In 2013, the Supreme Court of Colombia acquitted a journalist from defamation charges brought by a high-ranked public servant, ruling that the "principle of public relevance" established in its previous cases should be the guiding standard for deciding such cases. The Court noted that this principle allowed for a prioritization of the freedom of expression versus the person's right to reputation and noted that the status of the person and the content of the information would need to be considered while making such a determination.[790]

To similar effect, in 2014, the Supreme Court of Mexico was asked to rule on a petition related to mass emails that had criticized the actions of a state university's academic coordinator. The Court and dismissed the petition, holding that "the limits of criticism are broader if it concerns individuals who, because they are involved in public activities or because of the role they play in a democratic society, are exposed to a more rigorous oversight of their activities and statements than those private citizens who have no public influence."[791]

[787] *Lingens v. Austria*, App. No. 9815/82 (ECtHR 1986).
[788] *Eon v. France*, App. No. 26118/10 (ECtHR, Mar. 13, 2013).
[789] *Case of Herrera-Ulloa v. Costa Rica*, (Inter-American Court of Human Rights, July 2, 2004). www.corteidh.or.cr/docs/casos/articulos/seriec_107_ing.pdf
[790] *Gonzalez v. Serrano*, Rad. No. 38.909 (Corte Suprema de Justicia de Colombia, July 10, 2013).
[791] *Amparo Directo en Revisión*, 3123/2013 (Primera Sala de la Suprema Corte de Justicia de la Nación de México (SCJN), Feb. 7, 2014. Available at: http://www2.scjn.gob.mx/ConsultaTematica/PaginasPub/DetallePub.aspx?AsuntoID=156633&SinBotonRegresar

Also, in 2014, the Constitutional Court of Panama upheld the constitutionality of a law that had partially decriminalized insults to the honor of high-ranking public officials on the ground that the latter are subject to more oversight and scrutiny because of their positions and "public relevance." The court also noted that such officials should exhibit greater levels of tolerance with respect to speech against them.[792]

A similarly firm stance was taken by the Constitutional Court of the Dominican Republic in a 2016 case challenging the constitutionality of provisions of the Law on the Expression and Dissemination of Thought. The impugned provisions criminalized speech about public servants and was brought by newspapers within the country. In a landmark decision, the Court ruled that these provisions were unconstitutional since they violated freedom of expression and press freedoms, as well as preventing citizen oversight of the government.[793]

Several cases from Brazil have expressly noted that the *desacato* laws targeting speech against public servants are incompatible with the jurisprudence of the Inter-American system and have struck down criminal sentences awarded on this basis.[794]

This position has also been echoed in the laws of some countries. For instance, Argentina has repealed laws on libel and slander on issues of public interest and Nepal has promulgated a law stating that criticism of public figures is not tantamount to defamation.

Some jurisdictions do not apply the position espoused under international human rights law with respect to the idea that civil servants must tolerate more critical speech. Contrary to the position that public officials should be subject to greater scrutiny, punishment increases when the defamatory statement concerns a public figure or public official. This applies in 9 OSCE states:

792 *Advertencia de Inconstitucionalidad*, Expediente No. 478-08, at 749-766 (Órgano Judicial de la República de Panamá, Apr. 11, 2014).
793 *Acción directa de inconstitucionalidad*, Expediente No. TC-01-2013-0009, Judgment TC/0075/16 (Tribunal Constitucional de República Dominicana, Apr. 4, 2016).
794 Recurso Especial No. 1.640.084 - SP 2016/0032106-0 (Superior Tribunal de Justiça do Brasil, Dec. 15, 2016);. Juizado Especial Criminal Adjunto a 2da Vara Criminal da Comarca de Belford Roxo, Processo No. 0013156 - 07.2015.8.19.0008 (Tribunal de Justiça do Estado do Rio de Janeiro Comarca de Belford Roxo, July 4, 2016); Criminal da Comarca da Capital de Santa Catarina. Processo No. 0067370-64.2012.8.24.0023 (Criminal Court of the District of the Capital of Santa Catarina, March 17, 2015).

Andorra,[795] Bulgaria,[796] France,[797] Germany,[798] Italy,[799] Monaco,[800] the Netherlands,[801] Portugal,[802] and Turkey.[803] These states' laws provide that "defamation and/or insult committed against a public official carries a harsher punishment than the same act committed against a private person."[804] Imprisonment is a possible punishment in each case except for Bulgaria and France. Other countries punish defamation of public officials equally as harshly as other defamation cases. These provisions directly contradict the principles established by international courts and the right to freedom of expression as discussed above.

The specific regional and national approaches will be discussed in the section on national and regional legislative approaches to online defamation.

4. Heads of State

Particular issues and challenges arise with respect to speech that is critical of a head of state. Many states offer special protection from insult to heads of state, including Andorra, Azerbaijan,[805] Belarus,[806] Belgium, Denmark, Germany,[807] Greece, Iceland,[808] Italy, Kazakhstan, Malta, Monaco, the Netherlands, Poland,[809] Portugal, San Marino,[810] Slovenia, Spain, Sweden,

795 Andorran Criminal Code, Art. 173.
796 Bulgarian Criminal Code, Art. 148.
797 Lib Presse, Art. 32.
798 German Criminal Code, Art. 188.
799 Italian Criminal Code, Art. 595.
800 Monaco PFE, Art. 23, 25.
801 Dutch Criminal Code, Art. 267.
802 Portuguese Criminal Code, Art. 184.
803 Turkish Criminal Code, Art. 3.
804 Griffen et al, "Defamation and Insult Laws in the OSCE Region: A Comparative Study," Organization for Security and Co-operation in Europe, International Press Institute, March 2017.
805 Azeri Criminal Code, Art. 147 (Libel penalised with up to three years in prison), Art. 323 (Discrediting or humiliating the honour and dignity of the head of state penalised with up to five years in prison).
806 Belarus Criminal Code, Art. 188, (Libel penalised with up to three years in prison).
807 Note that those convicted of "disparaging the German president" under German Criminal Code, Art. 90 may also be stripped of certain political rights.
808 Icelandic Criminal Code, Arts. 235, 236 (Defamation and slander penalised with one and two years in prison, respectively); Icelandic Criminal Code, Art. 101 (Penalties for defamation and slander can be doubled if the victim is the president).
809 Criminal Code, Arts. 212, 216 (Defamation and insult penalised with up to one year in prison); Criminal Code, Art. 135(2) (Publicly insulting the president penalised with up to three years in prison).
810 Criminal Code, Art. 185 (Aggravated defamation penalised with up to imprisonment of the first degree); Criminal Code Art. 342 (Offence to the Captains Regent penalised with up to imprisonment of the third degree).

Tajikistan,[811] Turkey, Turkmenistan, Uzbekistan and Vatican City. In all these states, insult or defamation of a head of state is punishable with imprisonment. However, the application of these laws and number of cases varies dramatically. Turkey has one of the highest uses[812] and the country's presidential insult law grew at a staggering pace under President Erdogan.[813] Another interesting example is Tajikistan, which had no generally applicable criminal defamation laws, [814] but nonetheless criminalized public insult to its head of state: "Public insult of the Founder of Peace and National Unity – Leader of the Nation or slander against him,"[815] an offense carrying a penalty of 2 to 5 years in prison.

5. Religious Defamation

The issue of whether religious defamation, also known as blasphemy laws, should be criminalized has emerged as an important issue in human rights law.

While human rights law protects the freedom of religion, there have been increasing calls by some countries to criminalize blasphemy and Islamophobia. In recent years, this has been a particular demand by the Organization of Islamic Cooperation (OIC), which has argued that such speech constitutes a form of religious discrimination and hate speech.[816]

However, other countries have noted that this framing is often used to justify censorship and harsh punishments in a manner that is incongruous with international human rights law. In 2011, a resolution adopted by the Human Rights Council appeared to reconcile multiple competing positions by condemning discrimination on the basis of religion (that is, it protected *persons* from

811 Tajik Criminal Code, Art. 137.
812 Between August 2014 and March 2016 alone, 1,845 cases were reported to have been filed under this law, Turkish Criminal Code, Art. 299.
813 Griffen et al, "Defamation and Insult Laws in the OSCE Region: A Comparative Study."
814 Tajik Criminal Code, Art. 137.
815 Griffen et al, "Defamation and Insult Laws in the OSCE Region: A Comparative Study."
816 Marie Juul Petersen and Heini í Skorini, "Freedom of expression vs. defamation of religions: Protecting individuals or protecting religions?" London School of Economics and Political Science, Mar. 1, 2017. Available at: https://blogs.lse.ac.uk/religionglobalsociety/2017/03/freedom-of-expression-vs-defamation-of-religions-protecting-individuals-or-protecting-religions/

discrimination*)* without making any mention of criminalizing criticism of religious ideas or practices.[817]

Thus, while human rights law arguably protects an individual from false speech, it recognizes the limits of attempting to apply such rights to religion. Essentially, the veracity of statements offending religions can never be proved or disproved since they are fundamentally assertions about *ideas* without an objective "truth." Thus, "no faith-based belief can empirically be proven true or false."[818]

The OSCE, the UN, as well as the Inter-American and African human rights systems, under the *Joint Declaration on Defamation of Religions, and Anti-Terrorism and Anti-Extremism*[819] agreed that the concept of "defamation of religions" is in conflict with international standards, as religions do not have their reputations in the way that individuals do. As a result, the declaration provides that restrictions on free expression "should never be used to protect particular institutions, or abstract notions, concepts or beliefs, including religious ones."[820]

For its part, the UN Human Rights Committee has interpreted the ICCPR to establish that "[p]rohibitions of displays of lack of respect for a religion or other belief system, including blasphemy laws," generally violate the ICCPR. The committee noted exceptions for "advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence" under Article 20, paragraph 2. Still, the committee affirmed that even prohibitions under these exceptions must nevertheless comply with the strict requirements of provision by law,

817 *Combating intolerance, negative stereotyping and stigmatization of, and discrimination, incitement to violence and violence against, persons based on religion or belief*, U.N. Human Rights Committee, A/HRC/RES/16/18, Apr. 12, 2011.

818 Asma Uddin and Haris Tarin, *Rethinking the "Red Line": The Intersection of Free Speech, Religious Freedom, and Social Change, The Brookings Project on U.S. Relations with the Islamic World*, Nov. 2013. Available at: https://www.brookings.edu/wp-content/uploads/2016/06/Free-Speech_English_Web.pdf; "Combating Defamation of Religions," Issues Brief, Becket Fund for Religious Liberty, Oct. 29, 2009.

819 *Joint Declaration on Defamation of Religions, and Anti-Terrorism and Anti-Extremism Legislation*, the UN Special Rapporteur on Freedom of Opinion and Expression, the OSCE Representative on Freedom of the Media, the OAS Special Rapporteur and the African Commission Special Rapporteur on Freedom of Expression and Access to Information, Dec. 15, 2008. Available at: https://www.osce.org/fom/35639

820 *Tenth Anniversary Joint Declaration: Ten Key Challenges to Freedom of Expression in the Next Decade*, U.N. Special Rapporteur on Freedom of Opinion and Expression, the OSCE Representative on Freedom of the Media, the OAS Special Rapporteur on Freedom of Expression, & the ACHPR Special Rapporteur on Freedom of Expression and Access to Information, Feb. 3, 2010.

legitimate aim, and necessity and proportionality under Article 19. The committee determined that, as a result, using blasphemy laws "to discriminate in favour of or against" any particular belief system, or adherents of one religion over another, "or religious believers over non-believers," would be impermissible. In addition, using these prohibitions "to prevent or punish criticism of religious leaders or commentary on religious doctrine and tenets of faith" would also violate free expression.[821]

Lastly, the Rabat Plan of Action points out that the right to freedom of religion or belief "does not include the right to have a religion or a belief that is free from criticism or ridicule." The plan suggests that national blasphemy laws are counterproductive as they may chill all inter-and intra-religious dialogue and debate, as well as criticism, which itself may be "constructive, healthy and needed." The plan also alleges that many national blasphemy laws "afford different levels of protection to different religions and have often proved to be applied in a discriminatory manner."[822]

Despite these views from IHRL, laws relating to blasphemy have been a persistent feature in several countries with speech deemed to be offending religions severely penalized. Many countries still maintain criminal blasphemy laws with imprisonment as a possible punishment. Those countries include Algeria, Andorra, Austria, Bangladesh, Canada, Cyprus, Denmark, Finland, Egypt, El Salvador, Ethiopia, Gambia, Germany, Greece, Guyana, Ireland, India, Indonesia, Iraq, Israel, Italy, Jordan, Kazakhstan, Kuwait, Lebanon, Liechtenstein, Malaysia, Morocco, Oman, Pakistan, Poland, Portugal, the Russian Federation, Rwanda, San Marino, Spain, Sudan, Suriname, Switzerland, Tanzania, Thailand, Tunisia, Turkey, the United Kingdom (N. Ireland and Scotland only), Vatican City, and Western Sahara. Even more severe punishments persist in certain countries where blasphemy is punishable by death. These countries are: Bahrain, Brunei Darussalam, Iran, Nigeria, Pakistan, Qatar, Saudi Arabia and Somalia.

Perhaps surprisingly, human rights courts have in some cases also granted primacy to religious rights over free speech concerns. Two cases from the ECtHR are instructive - both on Article 188 of Austrian law which criminalizes the public disparagement of religion - notably,

821 *General comment no. 34,* UN Human Rights Committee (HRC).
822 "Rabat Plan of Action," *Annual Report of the United Nations High Commissioner for Human Rights*, A/HRC/22/17/Add.4, Jan. 11, 2013.

however, the maximum penalty is six months imprisonment or a fine.[823] In both cases, the court held that conviction under Article 188 (in one case for insulting Roman Catholic faith and in the other for insulting Islam) does not violate the defendant's free speech rights under the ECHR.[824] Notably, in the latter case, even as the Court affirmed its support for previous jurisprudence that held that the right to free speech extended included the right to "offend, shock, and disturb", the Court still held that "gratuitously offensive" statements would not be protected and that "expressions that seek to spread, incite or justify hatred based on intolerance, including religious intolerance, do not enjoy the protection afforded by Article 10 of the Convention"[825]

However, it is notable that some countries have taken a different stance. Some countries, including the Netherlands (2014), Iceland (2015), Malta (2016), and France (Alsace-Moselle, 2017), have recently abolished blasphemy laws. And in 2013, the Fourth Civil Chamber of Private Law of the Court of Justice of Sao Paulo decided that a video criticizing Islam constituted protected speech, reasoning that religious criticism was an aspect of the right to freedom of thought and expression.[826]

C. Regional Legislative Trends

1. Europe

Many countries in Europe maintain criminal defamation laws, particularly in Southern Europe (especially Greece and Italy, Portugal and Turkey), Central Europe (especially Hungary), Central Asia and Azerbaijan.[827] Occasional convictions of journalists also continue to take place

[823] Austrian Criminal Code, Art. 188: "Whoever, in circumstances where his or her behaviour is likely to arouse justified indignation, publicly disparages or insults a person who, or an object which, is an object of veneration of a church or religious community established within the country, or a dogma, a lawful custom or a lawful institution of such a church or religious community, shall be liable to up to six months' imprisonment or a day-fine for a period of up to 360 days."

[824] *Case of Otto-Preminger-Institut v. Austria*, App. No. 13470/87 (ECtHR, Sept. 20, 1994); *E.S. v. Austria*, App. No. 38450/12 (ECtHR, Oct. 25, 2018).

[825] *E.S. v. Austria*, App. No. 38450/12 (ECtHR, Oct. 25, 2018).

[826] Case No. 0192984- 85.2012.8.26.0100 (Tribunal de Justiça do Estado de São Paulo, Sept. 19, 2013). Available at: https://esaj.tjsp.jus.br/cposg/open.do

[827] "Azerbaijani journalist given two-year prison term," International Press Institute, Mar. 3, 2017. Available at: https://ipi.media/azerbaijani-journalist-given-two-year-prison-term/

in states typically considered strong defenders of media freedom such as Denmark, Germany and Switzerland.

In terms of objective components, the criminal codes of many OSCE states differentiate between defamation consisting of allegations of a particular fact and insult consisting of offensive expression. Accordingly, two separate provisions on 'defamation' and 'insult' are frequently provided (e.g., Belarus, Bulgaria, France). Notably, in Europe, criminal defamation provisions in the insult category commonly do not explicitly require that speech in question be false. A number of states expand this basic structure to include a third offence that covers defamation in which the speaker knows the fact to be false (e.g., Germany, Greece, Switzerland).[828]

Italy and Greece issue prison sentences for defamation, despite often suspending them and converting them into criminal fines.[829] In Germany, for instance, slander committed through the media is punishable with up to 5 years in prison. In Slovakia, defamation that causes "large-scale damage," e.g. loss of employment or divorce, offenders face up to eight years behind bars. Under certain qualifying circumstances, those convicted of the offence of "false accusation" in Portugal face up to eight years in prison. [830]

2. Americas

Some countries in the Americas criminalize defamation. Canada provides for up to five years in prison for defamatory libel known by the speaker to be false. Brazil also enshrines criminal defamation laws in its criminal code: the crime of *injuria*, concerning offenses to a person´s "dignity or decorum;" the crime of *difamação*, defined as an accusation that harms a person's reputation; and the crime of *calúnia*, defined as falsely accusing someone of a crime. [831]

However, states in the Americas have been moving away from criminalizing defamation. "Desacato" criminal laws have been repealed in Bolivia, Chile, Costa Rica, Honduras, Nicaragua,

828 Griffin et al, "Defamation and Insult Laws in the OSCE Region: A Comparative Study."
829 *See Belpietro v. Italy*, App. No. 43612/10 (ECtHR, Sept. 24, 2013); *Mika v. Greece*, App. No. 10347/10 (ECtHR, Dec. 19, 2013).
830 Portuguese Criminal Code, Art. 365
831 "Justiça Federal condena humorista por injúria contra deputada federal," Seção Judiciária de São Paulo, 2019. Available at: http://www.jfsp.jus.br/comunicacao-publica/indice-noticias/noticias-2019/10042019-justica-federal-condena-humorista-por-injuria-contra-deputada-federal/

Panama, Paraguay, Peru, and Uruguay.[832] Jamaica has repealed all its criminal defamation laws. Mexico has decriminalized federal defamation laws, and Argentina has repealed laws on libel and slander on issues of public interest.

The United States is quite unique in its laws to protect speech at the national level, severely limiting even civil actions for defamation. In 1964, the U.S. Supreme Court set a high bar for any civil libel action in *New York Times v. Sullivan*, declaring that public officials, in order to sue under defamation, must prove that the defendant's statement was false and made with "actual malice," i.e. "with knowledge that it was false or with reckless disregard of its truth or falsity."[833] The court went on to strike down a criminal defamation that lacked the "actual malice" standard in *Garrison v. Louisiana*, and find the common law crime of libel to be impermissibly vague in *Ashton v. Kentucky*. In the following decades, many American states repealed or watered down their criminal libel laws; fifteen out of fifty still had criminal libel statutes as of 2015, but they are rarely enforced.[834] The U.S. approach is, for the most part, not replicated elsewhere in the Americas.

3. Middle East

A number of Middle Eastern countries have and vigorously enforce criminal defamation laws, including Saudi Arabia, Israel, Algeria, and Turkey.

Article 26 of the Arab Charter on Human Rights guarantees "freedom of thought, conscience and opinion,"[835] but it does not take an official stance on defamation. Additionally, the charter provides for freedom of opinion and thought only insofar as it does not violate national laws.[836]

832 *Chapter IV - Laws on Contempt, Compulsory Membership, and Murder of Journalists*, Special Rapporteurship For Freedom Of Expression, Organization of American States, 2020.
833 *New York Times v. Sullivan*, 376 U.S. 254 (U.S. Supreme Court 1964).
834 A. Jay Wagner and Anthony Fargo, *Criminal Libel in the Land of the First Amendment*, Special Report for the International Press Institute, Sept. 2015. Available at: https://ipi.media/wp-content/uploads/2017/02/IPI-CriminalLibel-UnitedStates.pdf
835 Arab Charter on Human Rights, Art. 30.
836 Arab Charter on Human Rights, Article 27: "Persons from all religions have the right to practice their faith. They also have the right to manifest their opinions through worship, practice or teaching without jeopardizing the rights of others. No restrictions of the exercise of the freedom of thought, conscience and opinion can be imposed except through what is prescribed by law."

In a 2015 case, a Saudi writer was arrested for defaming a former ruler of the country. A 2014 counterterrorism law prohibits actions that "threaten Saudi Arabia's unity, disturb public order, or defame the reputation of the state or the king" as acts of terrorism. Under the law, authorities can hold and question a suspect for 90 days without a lawyer present.[837]

In Israel, defamation can constitute either civil or criminal offense and is punishable with up to one year in prison.[838] In April 2014, Israeli politician MK Revital Swid announced a new bill that would update the 1965 Defamation (Prohibition) Law to add provisions for Internet and digital communications. [839]

The Algerian Penal Code provides for the criminal punishment of all offences of defamation (article 296) and insult (article 297).[840]

In Turkey, human rights monitoring has focused on the abuse of provisions protecting the president and other public officials. Article 22 of the Constitution of the Republic of Turkey, which provides that "everyone has the right to freedom of communication, and secrecy of communication is fundamental." However, official data show 58,201 convictions in 2015 alone under the country's general insult law.

4. Africa

In Africa, Nigeria and Uganda have criminalized defamation, while South Africa has not.

In Nigeria, defamation can be treated as a tort or as a crime.[841] The 2015 Cyber Crime Act provides for imprisonment of up to three years for sending an online or electronic message that he

837 Linah Alsaafin, "Saudi writer arrested for insulting long-dead king," July 15, 2015. Available at: https://www.middleeasteye.net/news/saudi-writer-arrested-insulting-long-dead-king. *See also* Wafa Ben Hassine, *The Crime of Speech: How Arab Governments Use the Law to Silence Expression Online*, Electronic Frontier Foundation, 2016. Available at: https://www.eff.org/pages/crime-speech-how-arab-governments-use-law-silence-expression-online

838 Israeli Law 5725-1965, Sec. 5: "Any person who, with intent to injure, publishes a defamation to two or more persons other than the injured person, will be liable to one year's imprisonment, while publication of a defamation to one or more persons other than the injured person, constitutes a civil wrong."

839 Lidar Grave-Lazi, "New bill seeks to criminalize defamation in digital world," *The Jerusalem Post*, Apr. 2, 2015. Available at: https://www.jpost.com/Israel-News/New-bill-seeks-to-criminalize-defamation-in-digital-world-395984

840 *B. L. v. T. M.*, Decision No. 10874/11 (Algerian Court of Constantine, May 31, 2011).

841 Nigerian Criminal Code, Section 373.

knows to be false for an improper purpose, such as "causing annoyance, inconvenience danger, obstruction, insult, injury, criminal intimidation, enmity, hatred, ill will or needless anxiety."[842] In Nigeria, the only defenses to defamation are that the publication, at the time it is made, was for the public benefit or was true, or that the publication was privileged and made only to persons entitled to receive it.[843]

South Africa does not have a clear legislative framework regulating online defamation and courts have dealt with online defamation on a case-by-case basis. One proposed bill does make distributing or broadcasting a message harmful to another a criminal offense.[844] In the civil case of *Isparta v Richter 2013 6 SA 4529 (GP)*, a plaintiff sued defendants for defamation because the first defendant made comments on her Facebook profile and "tagged" the second defendant.[845]

Uganda provides for up to two years of imprisonment for criminal defamation.[846] Additionally, the Ugandan Computer Misuse Act of 2011 provides for imprisonment for up to one year for anyone who "willfully and repeatedly uses electronic communication to disturb or attempts to disturb the peace, quiet or right of privacy of any person with no purpose of legitimate communication."[847]

In one instructive case, the OHCR's Working Group on Arbitrary Detention found that an application of this law violated free expression. Ugandan authorities prosecuted human rights activist Stella Nyanzi under charges of offensive communication and cyberharassment as a result of her public Facebook posts. In these posts, she called the Ugandan president "a pair of buttocks" and criticizing the First Lady and Minister of Education for reneging on her campaign promise of

842 Nigerian Cybercrime Act, 2015.
843 *V. M. Iloabachie, Esq. v. Benedict N. Iloabachie*, Suit No: SC.137/2000 (Supreme Court of Nigeria, May 27, 2005).
844 Desan Iyer, "An Analytical Look Into The Concept Of Online Defamation In South Africa," Speculum Jurs, 2:32 (2018). Available at: http://specjuris.ufh.ac.za/sites/default/files/Iyer_proof_03.pdf
845 A. Roos and M. Slabbert, "Defamation on Facebook: Isparta v Richter 2013 6 SA 529," *PER: Potchefstroomse Elektroniese Regsblad*, 17:6 (2014), at 2845-2868. Available at: http://dx.doi.org/10.4314/pelj.v17i6.18
846 Ugandan Penal Code Act, Chapter 120, Sections 22 and 179. Available at: https://wipolex.wipo.int/en/text/170005
847 Ugandan Computer Misuse Act, 2011. Available at: https://ulii.org/ug/legislation/act/2015/2-6

providing sanitary napkins to schoolgirls.[848] The OHCR's Working Group on Arbitrary Detention found the state's measures excessive and disproportionate to the alleged offence, in violation of Article 19(3) of the ICCPR.[849]

5. Asia Pacific

Criminal defamation laws continue to be widely proliferated across the Asia Pacific Region. Lèse majésté laws are particularly common, and in some cases defamation laws have been updated to specifically include online defamation.

The criminal code of the People's Republic of China provides for punishing defamation with a prison sentence of up to 3 years.[850] In Japan, criminal defamation is punishable by imprisonment with or without work for up to 3 years, even if the statement in question was true.[851] In 2018, Cambodia made insulting any monarch punishable with up to one to 5 years in prison.[852]

The Thai Criminal Code punishes anyone who "defames, insults, or threatens" the royal family of Thailand with 3 to 5 years in prison.[853] In contrast, the Thai code provides for punishment of generic criminal defamation, for which truth is a defense if relating to public instead of "personal matters," with up to 1 year in prison or, if the defamation was by publication, 2 years in prison. [854]

[848] *Opinion No. 57/2017 concerning Stella Nyanzi* (Human Rights Council Working Group on Arbitrary Detention 2017). Available at: https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2017/04/A_HRC_WGAD_2017_57.pdf

[849] *Ibid.*

[850] People's Republic of China, Criminal Code, Art. 246: "Whoever, by violence or other methods, publicly humiliates another person or invent stories to defame him, if the circumstances are serious, shall be sentenced to fixed term imprisonment of not more than three years, criminal detention, public surveillance or deprivation of political rights."

[851] Japanese Criminal Code, Art. 230-1: "(1) A person who defames another by alleging facts in public shall, regardless of whether such facts are true or false, be punished by imprisonment with or without work for not more than 3 years or a fine of not more than 500,000 yen. (2) A person who defames a dead person shall not be punished unless such defamation is based on a falsehood."

[852] Prak Chan Thul, "Cambodia parliament adopts lese-majeste law, prompting rights concerns," *Reuters*, Feb. 14, 2018. Available at: https://www.reuters.com/article/us-cambodia-king/cambodia-parliament-adopts-lese-majeste-law-prompting-rights-concerns-idUSKCN1FY0RV

[853] Thai Criminal Code, Title XI, Chapter 3, § 112. Available at: http://library.siam-legal.com/thai-law/criminal-code-defamation-sections-326-333/

[854] *Ibid*, §§ 321–330.

Malaysia's Communications and Multimedia Act from 1998 states that anyone who insults state royalty online can face imprisonment or a fine. There have been multiple reports of criminal arrests for Facebook posts insulting Malaysian royal family.[855]

In India, the fundamental right to free speech (Article 19) is subject to "reasonable restrictions." Defamation, as outlined in Section 499 of the Indian Penal Code, is punishable by a fine or two years in prison. Section 499 has recently been extended to "electronic documents." Section 469 of the Indian Penal Code (forgery for purpose of harming reputation) has been amended by the Information Technology Act, 2000 to include committing forgery via electronic records. Offenders are liable for a statutory penalty of a find and/or up to 3 years imprisonment.[856]

D. Social Media Guidelines

Most major platforms, including Twitter, Facebook, and YouTube, rather than prohibiting defamation outright in their own community guidelines, address defamation by allowing users to submit local legal take-down requests. However, at least one platform—Snapchat—prohibits defamatory posts in its own community guidelines, but only when operating on the basis of a protected characteristic.

For instance, YouTube addresses defamation under its legal policies section. While YouTube generally defines defamation as "any untrue statement that is harmful to someone's reputation or causes someone to be shunned or avoided," YouTube's policy on defamation takes "local legal considerations" into account. However, declaring itself governed by U.S. law, YouTube explains that it is "not in a position to adjudicate the truthfulness of postings."[857] YouTube generally requires a court order before blocking or removing content.

Similarly, Twitter has no community guidelines banning defamation, but complies with legal demands to block local access to tweets determined to be illegal under local law, including

[855] Khairun-Nisaa Asari and Nazli Ismail Nawang, "A Comparative Legal Analysis of Online Defamation in Malaysia, Singapore and the United Kingdom," *International Journal of Cyber-Security and Digital Forensics*. 4:1 (2015)., at 314-326. DOI: 10.17781/P001548.
[856] Indian Penal Code, Sec. 469, 499.
[857] "Defamation," YouTube Help, as of Aug. 26, 2020. Available at: https://support.google.com/youtube/answer/6154230?hl=en

on the basis of defamation laws.[858] Facebook also allows the submission of defamation reports and asks filers if they have a court order to establish that the reported content is unlawful.[859] Likewise, WhatsApp's Terms of Service prohibit use of its services in "defamatory" ways.[860]

In contrast, Snapchat's community guidelines do mention defamation, albeit in a hate-speech related context. Its policy on hate speech and false information instructs users not to post content that "demeans, defames, or promotes discrimination or violence on the basis of" a protected characteristic.[861]

While the community guidelines of YouTube, Twitter, and Facebook do not address defamation itself, they do have policies on misinformation and disinformation that may sometimes overlap with defamation. These three platforms all ban posts of substantially manipulated media, fake accounts that impersonate others in misleading ways, misinformation or disinformation tending to suppress voting or census participation, and use of multiple accounts in ways that artificially manipulate conversations. These policies are discussed in more detail in the chapter on spreading false information.

E. Conclusion

The above sections have analyzed how restrictions on defamatory speech alternately comply and violate the international right to freedom of expression. International law guarantees the rights to reputation and privacy as well as free expression; the resulting tension has played out in both court cases and legislation, in the international and national spheres.

Our survey of international jurisprudence has noted four trends related to defamation. First, to be defamatory, statements must be factual rather than opinion-based. Second, despite state practice to the contrary, international courts consider criminalization of defamation too severe a form of restriction to comply with free expression. Third, free expression requires governments to

858 "Legal request FAQs," Twitter Help Center, as of Aug. 16, 2020. Available at: https://help.twitter.com/en/rules-and-policies/twitter-legal-faqs
859 Defamation reporting form, Facebook Help Centre, as of Aug. 16, 2020. Available at: https://www.facebook.com/help/contact/233704034440069
860 "WhatsApp Legal Info," WhatsApp, 2019. Available at: https://www.whatsapp.com/legal/.
861 "Community Guidelines," Snap Inc.

show particular restraint in restricting criticism of public figures and heads of state. Fourth, disagreement exists over the permissibility of criminal blasphemy laws protecting religions.

Our survey of national laws identified regional legislative trends in Europe, the Americas, the Middle East, Africa, and the Asia Pacific. Most defamation laws were originally passed to regulate offline conduct but have been interpreted to apply to online conduct as well. Others have been updated to specifically prohibit online defamation.

Many countries have criminal defamation laws. However, since 2000, over 30 states have taken steps towards doing away with criminalizing defamation, and over 10 countries have abolished the imposition of prison sentences for defamation.[862] These trends vary by region. Many European countries criminalize defamation. Asian Pacific countries have widely adopted criminal defamation laws, in addition to *lèse majésté* laws forbidding insult to royal families. A number of Middle Eastern countries have and vigorously enforce criminal defamation laws, including Saudi Arabia, Israel, Algeria, and Turkey. In Africa, Nigeria and Uganda have criminalized defamation, while South Africa has not. While some countries in the Americas criminalize defamation, such as Canada and Brazil, many have been repealing these laws; these include Bolivia, Chile, Costa Rica, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, and Uruguay.

[862] Amicus Curiae Brief of Intl. Human Rights Clinic at Yale Law School, *In re Emilio Palacio Urrutia et al*, No. P-143611.

VIII. CYBERHARASSMENT AND CYBERBULLYING

Many terms have been used to describe negative online behaviors and expressions that intentionally humiliate, annoy, attack, threaten, offend, or distress people, such as "cyberbullying," "cyberharassment," "cyber abuse," and "online harassment."[863] In this chapter, we use "cyberharassment." We define it as an umbrella term that encompasses harassment that takes place online, including behaviors such as cyberstalking, online harassment of minors (which we refer to as "cyberbullying"), the dissemination of revenge porn, and other harmful online conduct ("other cyber harms").

This chapter provides an overview of national and international law and jurisprudence related to some of the main forms of cyberharassment. First, this chapter briefly indicates which international human rights law (IHRL) treaties are relevant to cyberharassment as well as, for context, hate speech.

Second, this chapter surveys relevant principles of international jurisprudence, trends in national legislation, and notable court cases for each of the following categories of cyberharassment: cyberstalking, harassment of minors, sexual harassment, and other cyber harms including doxing. Most on-point court cases arose in national courts rather than international bodies.

Third, this chapter summarizes social media platform policies on hateful conduct. These platforms' community guidelines generally prohibit users from engaging in hate speech, threats of violence or property damage, cyberharassment or cyberbullying, and doxing of private personally identifiable information.

Finally, this chapter concludes with two key insights.

First, other than cyberbullying involving minors, cyberharassment activities tend to be at least "lightly criminalized," i.e. eligible for statutory penalties of less than two years in prison. The exception is online harassment involving minors, which is only specifically criminalized in a number of U.S. states; most countries have concluded cyberbullying is best handled within the

863 Marie-Helen Maras, *Cybercriminology*, Oxford: Oxford University Press, 2016.

education system. Since cyberharassment is a relatively recent phenomenon, a number of countries have yet to create laws that specifically address the issue. However, these countries generally have "offline" laws which can be applied to prosecute cyberharassment.

Second, our research identified very few cases in which courts found that prosecution of acts that constitute cyberharassment violated international or national rights of free expression. While rare, however, some court cases did find that criminal penalties for cases of cyberharassment violated free expression, generally when the speech at issue was politically-charged. However, this is an emerging legal area. Given the recent meteoric rise of social media use and the tendency of the law to lag behind technological changes, future international jurisprudence may identify more constraints on criminalizing cyberharassment.

A. Relevant IHRL Treaties

1. Hate Speech

Because the manifestation of hate speech online and the permissibility of its restriction is well-established, we focused our attention on other types of conduct, as requested: defamation and cyberharassment, including cyberbullying, cyberstalking, cyberharassment, and doxing. However, hate speech, and the consensus in international human rights law about frameworks for addressing state restrictions on hate speech, form an important backdrop for discussions of other forms of hateful conduct.

Multiple IHRL treaties actively obligate member states to prohibit certain types of hate speech. For instance, the ICCPR mandates in Article 20(2) that "[a]ny advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law."[864] The American Convention of Human Rights (ACHR) echoes this provision almost word for word but adds color and language as protected characteristics. Additionally, the International Convention on the Elimination of All Forms of Racial Discrimination (ICERD), in Article 4, commits member states to legally prohibit "all dissemination of ideas based on racial

[864] ICCPR, Art. 20.

superiority or hatred, incitement to racial discrimination, as well as all acts of violence or incitement to such acts against any race or group of persons of another colour or ethnic origin," as well as the financing or other forms of assistance to "racist activities."[865]

2. Cyber Harassment

No international or regional treaties *directly* address cyberharassment.[866] However, there are some provisions that address and encourage states to prohibit harassment and cyberharassment within particular domains, namely gender and the workplace.

Sexual and gender-based harassment falls under legally-binding human rights instruments on harassment against women and girls.[867] The African Union's Maputo Protocol to the ACHPR on the Rights of Women in Africa requires member states to take all appropriate measures to protect women and girls from sexual harassment in schools and the workplace, and punish perpetrators.[868] In the IACHR, states similarly agree to adopt legal measures to prevent sexual harassment, including in the workplace, educational institutions, and health facilities.[869]

Workplace harassment is governed by Convention 190 of the International Labor Organization, which prohibits "violence and harassment" in the world of work, defined as a "range of unacceptable behaviours and practices, or threats thereof, whether a single occurrence or repeated, that aim at, result in, or are likely to result in physical, psychological, sexual or economic harm, and includes gender-based violence and harassment."[870] The Committee on Economic, Social and Cultural Rights also stressed that the right to favorable and just working conditions

865 ICERD.

866 In this discussion of international human rights standards, we will exclude matters related to violence and discrimination (especially those that involved a protected group) since those often constitute hate speech, which is a more straightforward area in terms of prosecution and not in the scope of this section.

867 Defined as "any distinction, exclusion or restriction made on the basis of sex which has the effect or purpose of impairing or nullifying the recognition, enjoyment or exercise by women, irrespective of their marital status, on a basis of equality of men and women, of human rights and fundamental freedoms in the political, economic, social, cultural, civil or any other field." *Convention on the Elimination of All Forms of Discrimination against Women*, Art. 1., 1979.

868 *Protocol To The African Charter On Human And Peoples' Rights On The Rights Of Women In Africa*, African Union, 2003.

869 *Inter-American Convention On The Prevention, Punishment And Eradication Of Violence Against Women*, Organization of American States, 1994.

870 *Violence and Harassment Convention (No. 190)*, International Labour Organization, 2019.

(Article 7) implicitly includes freedom from physical and mental harassment.[871] The legal standards in these instruments on violence and harassment are considered violations of rights guaranteed specifically in the workplace, not generalizable to other spheres of life.

B. Cyberstalking

Cyberstalking generally entails repeated or persistent threatening conduct online with harmful effects for another. Numerous countries consider cyber stalking to be one of the more serious forms of harassment. This conduct is frequently criminalized in national legislation, although the harshness of the penalty varies across countries.

With the passage of the Council of Europe's Convention on Violence against Women (2013), Europe has witnessed a trend toward criminalizing stalking. Article 34 of the Convention requires signatory states to criminalize stalking, considered a serious form of harassment.[872] All but one of the EU member states have modified existing laws on harassment to accommodate this treaty obligation or established new laws prohibiting harassment and stalking.[873] While most of these laws do not explicitly address *cyber*stalking, countries are likely to apply laws criminalizing stalking offline to online conduct.

1. National Laws

a. Lightly Criminalized

Singapore and Nigeria both criminalize stalking with statutory penalties of fines or imprisonment for one year.

Singapore's Protection from Harassment Act (2014)[874] criminalizes "unlawful stalking" which involves stalking-like acts that intentionally cause, or are likely to cause, harassment, alarm

[871] *General Comment No. 23 on the right to just and favourable conditions of work (Article 7 of the ICESR)*, United Nations, Economic and Social Council, E/C.12/GC/23, 2016.
[872] "Parties shall take the necessary legislative or other measures to ensure that the intentional conduct of repeatedly engaging in threatening conduct directed at another person, causing her or him to fear for her or his safety, is criminalized." *Convention on Violence against Women*, Council of Europe, 2013.
[873] Suzan van der Aa, "New Trends in the Criminalization of Stalking in the EU Member States," *Eur. J. Crim. Policy Res*, 2018, 24:315–333.
[874] Singaporean Protection from Harassment Act, Chapter 256A. Available at: https://sso.agc.gov.sg/Act/PHA2014

or distress to the victim. The court is instructed to consider factors including frequency, circumstances, and impacts on victim health, safety, and reputation. Unlawful stalking is punishable by a fine not exceeding S$5,000 (€3,219)[875] or imprisonment for up to 1 year. The law gives the following circumstance as examples of unlawful stalking online: "Y repeatedly sends emails to Y's subordinate X with suggestive comments about X's body, or Y circulates revealing photographs of X to other classmates."

Nigeria's Cybercrime Act of 2015[876] considers an individual guilty of cyberstalking when he uses a public electronic communications network to "persistently send" messages that are either "grossly offensive or of an indecent, obscene or menacing character" or that "he knows to be false, for the purpose of causing annoyance, inconvenience or needless anxiety to another." The statutory penalty is a fine of at least ₦2,000,000 (€4,699)[877] or imprisonment for at least 1 year. Under "cyberstalking," Nigeria also groups behaviors involving cyber communications with intent to threaten, threats to cause physical harm, and threats to injure the reputation of the addressee with extortionary intent; the first two are discussed further in the section on other cyber harms.

b. Criminalized

Pakistan, India, and Uganda criminalize cyberstalking with statutory penalties of more than two years in prison.

Pakistan's Prevention of Electronic Crimes Act of 2016 explicitly establishes cyber stalking as a criminal offense. Under the law, a person commits cyber stalking when he (a) repeatedly uses an online mechanism to follow or attempt to contact someone "to foster personal interaction . . . despite a clear indication of disinterest by such person," (b) monitors someone's use of the Internet, email, or other electronic communication, (c) watches or spies on someone in a manner resulting "in fear of violence or serious alarm or distress" by the person, or (d) takes a photo or makes a video of someone and "displays or distributes it without his consent" in a harmful

875 1 SGD = 0.643884 EUR on May 23, 2020.
876 Nigerian Cybercrime Act of 2015, Section 24. Available at: http://www.nigerianlawguru.com/legislations/STATUTES/CYBERCRIME%20ACT%202015.pdf
877 1NGN = 0.002349EUR on May 23, 2020.

manner."[878] When the victim is an adult, the statutory penalties are a fine of up to 1,000,000 rupees (€5,119) or up to 3 years in prison. When the victim is an adult, the maximum penalties increase to 10,000,000 (€51,197) rupees or up to 5 years in prison.

In 2018, a defendant criminally prosecuted under the Pakistani provision received a six-year sentence for engaging in cyber-stalking and committing offenses against the dignity of a natural person through social media and email.[879] This legislation builds on the now-lapsed Prevention of Electronic Crimes Ordinances of 2007[880] and 2009,[881] which defined cyber stalking more broadly as using an online mechanism "to coerce, intimidate. or harass any person," imposing a fine of up to 300,000 rupees (€1,536) or a prison sentence of up to 7 years. If the victim is a minor, the penalty extends to up to 10 years.

India also criminalizes stalking, defined in a provision subtly different to Pakistan's, as either (a) repeatedly following and attempting to contact someone "to foster personal interaction . . . despite a clear indication of disinterest by such person" or (b) monitoring someone via an online or other electronic communication, or watching or spying on someone "in a manner that . . . interferes with the metal peace of such person."[882] This offense, added to the Indian Penal Code by the Criminal Law (Amendment) Act of 2013, specifies a statutory penalty of either 1 to 3 years in prison upon a first conviction or 3 to 5 years upon a second conviction.

Uganda's Computer Misuse Act (2011) establishes the crime of cyber stalking as any willful, malicious, and repeated use of electronic communication to harass another person and makes a threat with the intent to place that person in reasonable fear for his or her safety.[883]

[878] Pakistani Prevention of Electronic Crimes Act of 2016. Available at: http://www.na.gov.pk/uploads/documents/1470910659_707.pdf
[879] This case is described in more detail below. *Usman Bin Mehmood vs. the State & Another*, No. 77/2017 (Pakistan, Lahore Dist. Ct., 2018).
[880] Pakistani Prevention of Electronic Crimes Ordinances of 2007. Available at: https://www.ilo.org/wcmsp5/groups/public/---asia/---ro-bangkok/---sro-new_delhi/documents/genericdocument/wcms_300693.pdf
[881] Pakistani Prevention of Electronic Crimes Ordinances of 2009. Available at: http://nasirlawsite.com/laws/peco09.htm
[882] Indian Penal Code, Section 354D. Available at: https://devgan.in/ipc/section/354D/
[883] Ugandan Computer Misuse Act, 2011. Available at: https://ulii.org/ug/legislation/act/2015/2-6

Individuals guilty of cyber stalking are liable upon conviction to a fine or up to 5 years imprisonment.

2. Notable Case

Pakistan: Usman Bin Mehmood vs. the State & Another (2018)[884]

The defendant engaged in sexual relations with the plaintiff's wife, during the course of which he took intimate pictures and videos of her without her permission. He subsequently sent them via WhatsApp messages and fake email addresses for the purpose of blackmailing her. The charges were confirmed by forensic analysis of the defendant's phone and the email addresses. The defendant was sentenced to two years imprisonment and a fine of Rs. 200,000 (€ 1,142)[885] under s.20 of the Prevention of Electronic Crimes Act (offenses against the dignity of a natural person), 2 years imprisonment and a fine of Rs. 300,000 (€ 1,713)[886] under s.21 of PECA (offenses against the modesty of a natural person), and 2 years imprisonment and a fine of Rs. 200,000 (€ 1,142)[887] under s.24 of PECA (cyberstalking), with all sentences to run concurrently. Additionally, Rs. 10,000,000 (€ 57,100)[888] was awarded as compensation for damaging the social/private life of the victim as envisaged under s.45 of PECA (Order for payment of compensation).

C. Online Harassment of Minors ("Cyberbullying")

Although there is no clear universally accepted distinction between cyberharassment and cyberbullying, numerous countries have passed or amended laws to address cyberbullying in the context of the education system. Most laws targeting cyberbullying in the context of school-related bullying, involving victims and perpetrators who are minors or students, have been enacted in the past few years, often in reaction to high-profile tragic events involving young victims of cyberbullying. For instance, Italy passed Law No.71 in 2017 after a 14-year-old cyberbullying

884 *Usman Bin Mehmood vs. the State & Another*, No. 77/2017 (Pakistan, Lahore Dist. Ct., 2018). https://digitalrightsfoundation.pk/wp-content/uploads/2018/04/Cyber-Crime-FIR.pdf
885 1 PKR = 0.00571 EUR on May 23, 2020.
886 *Usman Bin Mehmood vs. the State & Another*.
887 *Ibid.*
888 *Ibid.*

victim jumped from the third floor of a building after a sexually explicit video of her that was filmed at a party began to circulate.[889]

Cyberbullying undertaken by minors is generally not criminalized, except in some US states. Instead, most of the laws below require educational institutions to deal with the issue, while noting that if the issue is serious enough (usually, if it violates the criminal code) the police may get involved. A European Parliament study has noted that criminalization is not common because many nations have concluded that a preventive and educational rather than punitive approach is more suitable when dealing with children, especially when many schools have the resources and structure required to take such an approach.[890] Furthermore, it is unclear as to whether criminalizing cyberbullying would result in extensive prosecution, especially considering the wide range of ages that are considered sufficient "for minimum responsibility" across jurisdictions.[891] The subsequent section provides a brief overview of national laws and the definitions of cyberbullying or bullying featured in those laws.

1. National Laws

a. Generally: Not Criminalized

South Korea's Act on the Prevention of and Countermeasures Against Violence in Schools (2008) establishes the responsibilities of state and local governments when dealing with school violence, along with setting out measures on how schools ought to address bullying (e.g. counseling, change of class). It defines cyberbullying as "constant or repeated actions" by which "students inflict emotional harm on other students by using the Internet, cell phones or other . . . devices" in order "to reveal personal information about a specific student or to spread lies or rumors about a specific student, and then inflict pain thereon."[892]

[889] Gavin Jones, "Italy passes law to fight cyber bullying," *Reuters*, May 17, 2017. Available at: https://www.reuters.com/article/us-italy-cyberbullying/italy-passes-law-to-fight-cyber-bullying-idUSKCN18D2GP
[890] *Cyberbullying Among Young People*, Policy Department C - Citizens' Rights and Constitutional Affairs, European Parliament, 2016. Available at: https://www.europarl.europa.eu/RegData/etudes/STUD/2016/571367/IPOL_STU(2016)571367_EN.pdf
[891] *Ibid.*
[892] South Korean Act on the Prevention of and Countermeasures Against Violence in Schools, 2008. Available at: http://elaw.klri.re.kr/eng_mobile/viewer.do?hseq=24031&type=new&key=

Japan's Anti-Bullying Law (2013) requires that schools report "serious cases" of bullying to appropriate governmental bodies and set up special panels to address the issue. The law defines serious cases as forms of bullying that cause serious physical and mental harm to children, even forcing them to be absent for extended periods of time. Along those lines, the law tasks central and local governments with monitoring online bullying activity and demands that schools involve the police if acts of bullying can be considered criminal (e.g. a threatening email which could constitute blackmail, calling someone a 'creep' which could be defamation). [893]

The Anti-Bullying Act (2013) of the Philippines is comparable to the Japanese law; it mandates that elementary and secondary schools adopt policies to address bullying, and that the principal or designated school officer notify law enforcement if they believe that criminal charges could be pursued against the perpetrator.[894] The Act places bullying in an educational context, defining it as a student's "severe or repeated use" of "written, verbal or electronic expression, or a physical act or gesture . . . directed at another student" which, as a result, places him "in reasonable fear of physical or emotional harm" or property damage, creates a hostile school environment for him, infringes on his rights at school, or "materially and substantially" disrupts the school's "educational process" or "orderly operation." The law specifies a non-exhaustive list of actions that will constitute bullying: unwanted physical contact (including fighting, tickling, and "inflicting school pranks"), an act that damages "a victim's psyche and/or emotional well-being," and slanderous statements or accusations, including foul language or name-calling directed at the victim or criticisms of the victim's "looks, clothes and body."

Chile's Law about School Violence (2011) amended the General Law on Education to set out various definitions, procedures, and penalties governing how schools address bullying. The law defines bullying as "repeated aggression or harassment, carried out outside or within the educational establishment by students who use "a situation of superiority" or the victim's "defenselessness" and cause the victim "abuse, humiliation or well-founded fear of being exposed

[893] "Diet passes anti-bullying legislation," *The Japan Times*, June 21, 2013. Available at: https://www.japantimes.co.jp/news/2013/06/21/national/upper-house-enacts-anti-bullying-law/
[894] Republic Act No. 10627, Congress of the Philippines Metro Manila, Fifteenth Congress, Third Regular Session. Available at: https://www.officialgazette.gov.ph/2013/09/12/republic-act-no-10627/

to a serious illness," whether via technological or other means. The law also notes that bullying can also involve with violence (physical or psychological) committed by other school community members like teachers. The law instructs schools to create committees and internal regulations to promote "peaceful coexistence," among other initiatives. [895]

In 2015, Brazil passed a similar law creating a national program to combat bullying, which is defined as an "act of physical or psychological violence that is intentional and repetitive, that occurs without evident reason, and is practiced . . . in order to intimidate or attack the victim, causing pain and distress," in which the involved parties are in a relationship of inequality of powers. Virtual bullying comprises "making disparaging remarks, sending messages invading the person's privacy, and/or sending or tampering with photos and personal data that result in suffering or that create psychological and social embarrassment." [896]

Italy's Law No. 71/2017 is similar to most of the laws mentioned above, as it requires schools to educate students on responsible internet use and designate staff to address cyberbullying. The law defines cyberbullying as any "form of psychological pressure, aggression, harassment, blackmail, injury, insult, denigration, defamation, identity theft, alteration, illicit acquisition, manipulation, unlawful processing of personal data of minors and/or dissemination made through electronic means" with the "intentional and predominant purpose" of isolating "a minor or a group of minors by putting into effect a serious abuse, a malicious attack or a widespread and organized ridicule." Though it does not criminalize cyberbullying, the law does note that certain criminal sanctions can be extended to cyberbullying (e.g. sanctions related to insults, defamation, threats, stalking). [897]

[895] Law No. 20536, Sobre Violencia Escolar, Chile, 2011. Available at: https://www.leychile.cl/Navegar?idNorma=1030087

[896] "Brazil: New Law Creates Program to Combat Bullying Nationwide," Global Legal Monitor, Library of Congress, Nov. 13, 2015. Available at: https://www.loc.gov/law/foreign-news/article/brazil-new-law-creates-program-to-combat-bullying-nationwide/

[897] Giovanni Ziccardi, *Cyber Law in Italy*, 3rd ed., Wolters Kluwer, Jan 14, 2020; Pietro Ferreara et. al., "Cyberbullying a modern form of bullying: let's talk about this health and social problem," *Italian Journal of Pediatrics*, 44:14 (2018). Available at: https://ijponline.biomedcentral.com/articles/10.1186/s13052-018-0446-4#ref-CR13

An interesting component of Law No.71/2017 is its involvement of data controllers: the law empowers underage victims who are at least 14 years old and their parents to contact the relevant data controller or a website provider to request the removal, blocking or takedown of content that constitutes cyberbullying. The recipients of the request must respond within 48 hours; otherwise, the complainants can take the issue to the Italian Data Protection Authority (the Garante).[898]

b. Notable Exception: Light Criminalization in the United States

There are some U.S. states that lightly criminalize cyberbullying involving minors, declaring it a misdemeanor punishable by one year in prison or less. For instance, Maryland's Misuse of Interactive Computer Service define, as a criminal offense, maliciously using an "interactive computer service" in a way that "inflicts serious emotional distress on a minor or places a minor in reasonable fear of death or serious bodily injury," and with the "intent (1) to kill, injure, harass, or cause serious emotional distress to the minor or (2) to place the minor in reasonable fear of death or serious bodily injury." Violators are guilty of a misdemeanor, punishable by imprisonment for up to one year and/or a $500 maximum fine."[899] Likewise, Louisiana's Act No. 989 criminalizes cyberbullying, defined as transmitting "any electronic textual, visual, written, or oral communication with the malicious and willful intent to coerce, abuse, torment, or intimidate a person under the age of eighteen." The offense is punishable by a fine of up to 500 USD or imprisonment of up to 6 months in jail.[900]

D. Sexual Harassment

One of the most common forms of sexual harassment that takes place online is "revenge porn," which generally entails the online distribution of someone's intimate images or videos without consent, often with the intent of causing embarrassment or distress. The online distribution

[898] "Italy: Cyberharassment, violations of privacy, cybercrime, actual violence," Council of Europe Portal, as of Aug. 26, 2020. Available at: https://www.coe.int/en/web/cybercrime/italy#{%2255438873%22:[]}
[899] Maryland, House Bill Chapter 369, Misuse of Interactive Computer Service, 2013. Available at: http://mgaleg.maryland.gov/2013RS/chapters_noln/Ch_369_hb0396T.pdf
[900] Louisiana, House Bill 1259, 2010.

of sexual images without consent and similar actions that fundamentally violate someone's privacy and bodily autonomy tend to be criminalized, though the harshness of the penalty varies across countries.

1. National Laws

a. Light Criminalization

Kenya, England, and France have all made "revenge porn" criminal offenses punishable by up to 2 years in prison. Kenya's Computer Misuse and Cybercrimes Bill (2018) criminalizes the wrongful distribution of obscene or intimate images of another person with a fine up to 200,000 shillings (€1,715)[901] and/or 2 years imprisonment.[902] In 2015, England and Wales made it a criminal offence punishable by up to 2 years in prison to disclose private sexual photographs or videos without the subject's consent, with the intent of causing that person distress.[903] A year later, France's Digital Republic Law criminalized the same act, making it punishable by up to 2 years in prison and/or a €60,000 fine.[904]

b. Criminalization

The Philippines has one of the earliest revenge porn laws, dating back to 2009. The Anti-Photo and Video Voyeurism Act of 2009 declares it a prohibited act to photo or video "a person or group of persons performing sexual act or any similar activity" or a person's "private area," such as "the naked or undergarment clad genitals, pubic area, buttocks or female breast," without obtaining "the consent of the person/s involved," and in "circumstances in which the person/s has/have a reasonable expectation of privacy" and to be complicit in its distribution. It is also prohibited to copy or reproduce, sell or distribute, or publish or broadcast, or electronically show

[901] 1 KES = 0.00857272 EUR on May 23, 2020.
[902] Kenya, Computer Misuse and Cybercrimes Act, 2018. Available at: http://kenyalaw.org/kl/fileadmin/pdfdownloads/Acts/ComputerMisuseandCybercrimesActNo5of2018.pdf
[903] UK, Criminal Justice and Courts Act, 2015. Available at: http://www.legislation.gov.uk/ukpga/2015/2/section/33/enacted
[904] France, Law No. 2016-1321, 2016. Available at: https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000033202746&categorieLien=id

or exhibit, such images. Doing so is punishable by a fine between P100,000 and P500,000 (€1,807 and €9,034)[905] and/or imprisonment between 3 and 7 years.[906]

Canada criminalized sharing someone's intimate images without consent in 2015.[907] An intimate image is defined as "an image that depicts a person engaged in explicit sexual activity or that depicts a sexual organ, anal region or breast," where "the person depicted had a reasonable expectation of privacy at the time of the recording and had not relinquished his or her privacy interest" by the time the images were shared. Those found guilty of this offence may face imprisonment up to 5 years, seizure of devices used to share the image, and an order to reimburse various costs incurred by the victim.

In Israel, it is a crime to distribute images or videos of someone without their consent that focus on their sexuality, including by means of photo-editing or video-editing, in a manner that facilitates identifying the person, and in a way that may degrade or shame the person. However, the law provides for a defense if the images contained a criticism of a public official in his or her official capacity and was a reasonable means of communicating that criticism. The offense is punishable by up to five years of imprisonment, and the perpetrator can be mandated to pay monetary compensation to the victim.[908]

South Africa is considering passing a similar law, the Cybercrimes and Cybersecurity Bill, which would criminalize making available, broadcasting, or distributing electronic messages that, among other things, distribute intimate images without consent. The law would make the act punishable by a fine and/or imprisonment not exceeding 3 years.[909]

[905] 1 PHP = 0.0180688 EUR on May 23, 2020.
[906] Republic Act No. 9995, An Act Defining and Penalizing the Crime of Photo and Video Voyeurism, Prescribing Penalties Therefor, and for Other Purposes, 2009, Congress of the Philippines. Available at: https://lawphil.net/statutes/repacts/ra2010/ra_9995_2010.html
[907] "What are the potential legal consequences of cyberbullying?" Public Safety Canada, last modified 2018. Available at: https://www.publicsafety.gc.ca/cnt/ntnl-scrt/cbr-scrt/cbrbllng/prnts/lgl-cnsqncs-en.aspx#a01
[908] Prevention of Sexual Harassment (PSH) Law 5758-1998, *Sefer Hahukim [Official Gazette]*, No. 1661 at 166, as amended; PSH Law (Amendment No. 10), 5774-2014; Ruth Levush, "Israel: Prohibition of Online Distribution of Sexual Images Without Consent," *Global Legal Monitor*, Jan. 10, 2014. Available at: http://www.loc.gov/law/foreign-news/article/israel-prohibition-of-online-distribution-of-sexual-images-without-consent/
[909] South Africa, Cybercrimes and Cybersecurity Bill, 2017. Available at: https://www.justice.gov.za/legislation/bills/CyberCrimesBill2017.pdf

In South Africa, victims of sexual harassment are able to take advantage of the Protection from Harassment Act (2011) which empowers complainants to apply to the court for a protection order against harassment. In order to grant such an order, the court must be satisfied that the respondent has engaged in harassment that is harming or has harmed the victim in a manner that will likely continue if a protection order is not issued immediately. If the harassment is taking place online, the court may demand that the electronic communications service provider turn over information about the respondent if his identity or address is unknown. Violating a protection order or making a false statement in an affidavit are punishable by a fine or imprisonment not exceeding 5 years. [910]

The South African act's definition of sexual harassment includes (a) unwelcome sexual attention from someone who does or should reasonably know it is unwelcome, (b) "unwelcome explicit or implicit behavior, suggestions, messages or remarks of a sexual nature" which offend, intimidate, or humiliate the victim, in circumstances where a reasonable person would have anticipated the victim would be offended, humiliated or intimidated, (c) an explicit or implicit "promise of reward for complying with a sexually oriented request," or (d) an explicit or implicit reprisal or threat of reprisal "for refusal to comply with a sexually oriented request."[911]

Pakistan's Prevention of Electronic Crimes Act of 2016 also effectively criminalized revenge porn by including, in the definition of cyber stalking, taking a photo or making a video of someone and displaying or distributing it without his consent in a harmful manner. When the victim is an adult, the statutory penalties are a fine of up to 1,000,000 rupees (€5,119) or up to 3 years in prison. When the victim is an adult, the maximum penalties increase to 10,000,000 (€51,197) rupees or up to 5 years in prison.[912]

2. Notable Cases

a. The United States: Austin v. Illinois (2019)

910 South Africa, Protection from Harassment Act 17, 2011, Government Gazette Vol. 558, No. 34818, Dec. 5, 2011. Available at: https://www.gov.za/documents/protection-harassment-act

911 *Ibid.*

912 Pakistan, Prevention of Electronic Crimes Act, 2016. Available at: http://www.na.gov.pk/uploads/documents/1470910659_707.pdf

It is worth noting that revenge porn is criminalized in almost every single state in the United States, despite the country's extreme commitment to freedom of speech as enshrined in the First Amendment of the U.S. Constitution.[913] Appellate courts in Illinois, Wisconsin, and Vermont have all upheld revenge porn laws against constitutional challenges on free speech grounds, under the same principles as laws prohibiting nonconsensual disclosures of sensitive financial or medical information.

In the facts of one notable case, *Austin v. Illinois*, defendant Bethany Austin broke up with her fiancé Matthew after discovering that he had been unfaithful, and sent a letter detailing her version of events to her and her fiancé's families, including naked pictures that Matthew had received from his mistress, the victim. When Matthew found out about the letters, he contacted the police. The victim emphasized that she had intended for the photos to be seen by *only* Matthew.

Section 11-23.5(b) of the Illinois Criminal Code makes it a class 4 felony, punishable by up to three years imprisonment and/or fine up to $25,000 (€22,934),[914] to disseminate private sexual images without consent. Austin was charged with violating this law, which she argued was an unconstitutional restriction on the right to free speech that does not serve a compelling government interest. The intermediate appellate court agreed with Austin, dismissing the charges and finding the law unconstitutional.

The Illinois Supreme Court reversed the decision in a 5-2 vote, finding (1) the Illinois law is a content-neutral restriction because it is based more on non-consensual procurement rather than nudity, (2) the speech is not of public concern, since it is fundamentally about the victim's private activities (i.e. private significance), and (3) the government has a substantial interest in protecting individual privacy. It also noted that a wrongful purpose is "inherent in the act of disseminating an intensely personal image without the consent of the person portrayed." The Court reasoned that, "[v]iewed as a privacy regulation," the law "is similar to laws prohibiting the unauthorized disclosure of other forms of private information, such as medical records, biometric data, or Social Security numbers," and remarked that "[t]he entire field of privacy law is based on the recognition

[913] "46 States + DC + One Territory Now Have Revenge Porn Laws," Cyber Civil Rights Initiative, as of Aug. 26, 2020. https://www.cybercivilrights.org/revenge-porn-laws/
[914] 1 USD = 0.917352 EUR on May 23, 2020.

that some types of information are more sensitive than others, the disclosure of which can and should be regulated."[915]

Similar revenge porn laws in have been upheld by appellate courts in Wisconsin and Vermont, which found that sexually explicit images deserved privacy in the same way as other sensitive information, like medical records and financial data.[916] As the Vermont Supreme Court reasoned: "[f]rom a constitutional perspective, it is hard to see a distinction between laws prohibiting nonconsensual disclosure of personal information comprising images of nudity and sexual conduct and . . . other categories of nonpublic personal information."[917] The court assessed that the government has a strong interest in protecting all of these from disclosure.

b. Kazakhstan: WA and WB v. Mamedov (2019) (Civil/Constitutional)

In the case of *WA and WB v. Mamedov*, the Supreme Court of Kazakhstan held a defendant civilly liable to two women for posting a sexual video of them without their consent. The ruling was based on Kazakh constitutional law rather than international human rights law.[918]

The defendant Mamedov E.A. had recorded the two women kissing at a movie theater without their consent. He posted this video online, criticized their sexual orientation and called for them to be outed and shamed. As a result, the applicants faced harassment and threats (including hate speech), eventually causing them to flee the country for multiple months.

The Supreme Court of Kazakhstan ruled in favor of the two women, finding that Mamedov had violated their right to privacy and caused them significant harm by posting the video on social media, thereby overturning the initial appellate court judgment that held that he was using his freedom of expression to defend societal morals.

915 *Illinois v. Austin*, 2019 IL 123910 (Supreme Court of Illinois 2019). Available at: https://cases.justia.com/illinois/supreme-court/2019-123910.pdf?ts=1571410923

916 Deanna Paul, "Is revenge porn protected speech? Lawyers weigh in, and hope for a Supreme Court ruling," *The Washington Post,* Dec. 26, 2019. Available at: https://www.washingtonpost.com/nation/2019/12/26/is-revenge-porn-protected-speech-supreme-court-may-soon-weigh/

917 *Vermont v. VanBuren*, 2018 VT 95 (Vermont Supreme Court 2017).

918 *WA and WB v. Mamedov*, 6001-19-00-3гп/389 (Supreme Court of Kazakhstan, July 30, 2019). *See* "WA and WB v. Mamedov," Columbia Global Freedom of Expression Database*,* July 30, 2019.

Specifically, the Supreme Court held that Mamedov had violated the right to privacy guaranteed by Article 18(1) of the Constitution and Article 115(3) of the Civil Code. The court required Mamedov to pay compensatory monetary damages to the women.

E. Other Cyber Harms (Annoyance, Alarm, Threat, Doxing, etc.)

An additional category of cyberharassment, that does not fall under cyberstalking, bullying of minors, or online sexual harassment, can be described broadly as a residual category of online conduct resulting in harm, whether mental, psychological, physical, or economic. While there is some variation in the types of harm and severity discussed across national jurisdictions, many national laws criminalize or lightly criminalize this type of behavior online.

Countries are most likely to criminalize conduct that facilitates violence or physical damage, which sometimes falls under incitement to violence. However, countries also impose criminal sanction for other harmful online behaviors that result in fear, alarm, annoyance, or distress. Many but not all countries require that the perpetrator possess an intent to cause harm or knowledge that their behavior would cause harm. The publication of identity information that can cause similar harm (i.e. doxing) is a particular subset of this conduct that has been criminalized in multiple countries.

1. National Laws

a. Not Criminalized

The South African Protection from Harassment Act (2011)[919] defines cyberharassment as "conduct that the respondent knows or ought to know causes mental, psychological, physical, or economic harm" or that "inspires the reasonable belief that harm may be caused to the complainant" by unreasonably engaging in verbal, electronic or other communications, or conduct that amounts to sexual harassment. However, the Act does not criminalize such conduct, instead providing for the issuance of civil protection orders on the victim's behalf.

[919] South Africa, Protection from Harassment Act 17, 2011.

b. Lightly Criminalized

The United Kingdom, Singapore, and many U.S. states lightly criminalize these other cyber harms, with statutory punishments of fines or sentences of imprisonment not to exceed 2 years.

The UK's Malicious Communications Act (1988, most recently revised in 2015)[920] states that sending messages or letters electronically that the government deems to be "of an indecent or grossly offensive nature," with the purpose of causing "distress or anxiety" to the recipient or other person to whom the sender intends the message to be passed can lead to a prison sentence of six months and a fine.

Singapore's Protection from Harassment Act (2014)[921] criminalizes intentional harassment, defined as (a) any act, i.e. using any words or behavior or making any communication that is "threatening, abusive or insulting words or behavior," or publishing the target's identity information, which (b) intentionally causes the victim "harassment, alarm, or distress." Identity information may include the target's name, address, phone, birthday, recordings or images. This offense is punishable by a fine up to S$5,000 (€3,219)[922] and/or imprisonment not exceeding 6 months.

The law provides examples of intentional online harassment. In the first, X is guilty of intentional harassment: "X and Y were formerly in a relationship which has since ended," and "X writes a post on a social media platform making abusive and insulting remarks about Y's alleged sexual promiscuity." In another post, "X includes Y's photographs and personal mobile number, intending to cause Y harassment by facilitating the identification or contacting of Y by others." Even though "Y did not see the posts . . . [Y] receives and is harassed by telephone calls and SMS messages from strangers (who have read the posts) propositioning Y for sex." In contrast, the following example does not rise to the level of intentional harassment: "X records a video of Y driving recklessly in a car on the road. X posts the video on an online forum, where people share

920 U.K., Malicious Communications Act, 1988. Available at: http://www.legislation.gov.uk/ukpga/1988/27/section/1/data.htm

921 Singapore, Protection from Harassment Act, Chapter 256A, revised 2015. Available at: https://sso.agc.gov.sg/Act/PHA2014

922 1 SGD = 0.643884 EUR on May 23, 2020.

snippets of dangerous acts of driving on the road. X posts the video with the intent to warn people to drive defensively."

Singapore's act also lightly criminalizes unintentional harassment, defined as using words or behavior or making communications that are "threatening, abusive words, or behaviour," that are perceived by the victim and "likely" to cause the victim harassment, alarm, or distress. This offense is punishable by a fine only, of up to S$5,000 (€3,219).[923] Plausible defenses include if the accused proves he had no reason to believe that his actions would be seen and/or that they were reasonable. The act gives the following example of such an offence: "X and Y are classmates. X posts a vulgar tirade against Y on a website accessible to all of their classmates. One of Y's classmates shows the message on the website to Y, and Y is distressed. X is guilty of an offence under this section."

The act also criminalizes communications, including doxing, that may provoke or facilitate violence. First, the act prohibits the use, against another person, of "any threatening, abusive or insulting words or behaviour, or make any threatening, abusive or insulting communication to another person" with the intent of provoking violence or making the victim fear they will be subject to violence. Second, the act prohibits publishing **"**identity information" with the same intent of provoking or making the victim fear violence. The provocation and facilitation of violence in such manners is punishable by a fine up to S$5,000 (€3,219)[924] and/or imprisonment not exceeding 12 months. The defenses mentioned in the above paragraph for unintentional harassment (reasonable actions, expectations of not being perceived) are also provided for.

The law provides two social-media related examples: one on threats and abuse, and the second on doxing. In the first, X is guilty: "X and Y are classmates. X writes a post with threatening and abusive remarks against Y on a website accessible to all their classmates. X writes a subsequent post on the same website, stating Y's identity information and stating 'Everyone, let's beat Y up!'" In the second, B is guilty: "X writes a post (on a social media platform to which Y

[923] Singapore, Protection from Harassment Act, Chapter 256A, revised 2015.
[924] *Ibid.*

does not have access) containing threats of violence against Y and calling others to 'hunt him down and teach him a lesson.' B posts Y's home address in reply to X's post."

Higher penalties (up to double the maximum punishment, so up to 2 years) are permitted for offences against "vulnerable persons," defined as individuals who are, "by reason of mental or physical infirmity, disability or incapacity, substantially unable to protect [themselves] from abuse, neglect or self-neglect." The Singaporean law also permits higher penalties for subsequent offences and instances where the offender had been in an "intimate relationship" with the victim.

Some US states lightly criminalize online conduct that the perpetrator knows is likely to cause psychological harm, declaring it a misdemeanor. For instance, the Code of Alabama establishes that communicating with someone in a manner "likely to harass or cause alarm" is punishable by up to 3 months imprisonment and/or a fine up to $500 (€459).[925]

California's Penal Code makes it a criminal offense to, with the purpose of imminently causing the victim "unwanted physical contact, injury, or harassment," electronically disseminate or make available "personal identifying information, including . . . a digital image of another person, or an electronic message of a harassing nature about another person, which would be likely to incite or produce that unlawful action." An offender under this provision is guilty of a misdemeanor punishable by a fine of up to $1,000 (€917)[926] or up to 1 year in a county jail.[927]

Under Delaware's Criminal Code, a person may be subject to one year of imprisonment and/or fines up to $2,300 if he "insults, taunts or challenges another person or engages in any other course of alarming or distressing conduct which serves no legitimate purpose," in a manner he "knows is likely to provoke a violent or disorderly response or cause a reasonable person to suffer fear, alarm, or distress."[928]

c. Criminalization

[925] Code of Alabama, Section 13A-11-8, Harassment or harassing communications, 1977. Available at: http://alisondb.legislature.state.al.us/alison/codeofalabama/1975/13A-11-8.htm
[926] 1 USD = 0.917342 EUR on May 23, 2020.
[927] California Penal Code, Title 15, §§ 626–653.75. Available at: https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?lawCode=PEN&sectionNum=653.2
[928] Delaware Code, Title 11, § 1311, Harassment. Available at: https://codes.findlaw.com/de/title-11-crimes-and-criminal-procedure/de-code-sect-11-1311.html

Though South Africa does not currently criminalize generic online harassment causing harm, the South African legislature has proposed a Cybercrimes and Cybersecurity Bill that would criminalize making available, broadcasting, or distributing (by any means of a computer system) electronic messages that meet any of the following criteria: inciting or threatening violence or property damage; intimidating, encouraging or harassing someone to harm himself or someone else; are "inherently false and aimed at causing mental, psychological or economic harm;" or distribute "intimate images without consent."[929] Such conduct would be punishable by a fine and/or imprisonment not exceeding three years.[930]

Kenya's Computer and Cybercrimes Bill (2017)[931] declares it an offense for someone, who knows or ought to know his conduct is likely to detrimentally affect another or cause them to fear violence or property damage, to "willfully and repeatedly communicate[ ], either directly or indirectly, with" that person or anyone known to them. This offence is punishable by up to 10 years imprisonment and/or 20 million shillings (€171,439).[932] The cyberharassment provision in Kenya's Computer Misuse and Cybercrimes Bill (2018)[933] is the same as the cyberstalking and cyberbullying provision of the earlier bill, though it also criminalizes conduct that is "in whole or in part, of an indecent or grossly offensive nature and affects the person." Moreover, the bill criminalizes the wrongful distribution of obscene or intimate images of another person with a fine up to 200,000 shillings (€1,715)[934] and/or two years imprisonment.

Nigeria's Cybercrime Act (2015) similarly criminalizes transmitting any communication, through information and communication technologies, with intent to bully, threaten or harass another person, where such communication places another person in fear of death, violence or personal bodily injury or to another person.[935] Such offenses are subject to imprisonment for a

[929] On June 23, 2020, the Select Committee on Security and Justice in the National Council of Provinces approved amendments to the bill. Cybercrimes Bill, Ellipsis. Available at: https://www.ellipsis.co.za/cybercrimes-bill/
[930] South Africa, Cybercrimes and Cybersecurity Bill, 2017. Available at: https://www.justice.gov.za/legislation/bills/CyberCrimesBill2017.pdf
[931] Kenya, Computer and Cybercrimes Bill, 2017. Available at: http://kenyalaw.org/kl/fileadmin/pdfdownloads/bills/2017/ComputerandCybercrimesBill_2017.pdf
[932] 1 KES = 0.00857272 EUR on May 23, 2020.
[933] Kenya, Computer Misuse and Cybercrimes Act, 2018.
[934] 1 KES = 0.00857272 EUR on May 23, 2020.
[935] Nigeria, Cybercrime Act, 2015.

term of not less than 10 years and/or a fine of not less than N25,000,000 (€58,739).[936] Interestingly, by way of comparison, the Act prosecutes online threats to kidnap or injure another, or injure the property or reputation of an addressee with extortionary intent, subject to imprisonment for a term of not less than 5 years or a fine of not less than N15,000,000 (€35,238).[937]

Under Uganda's Computer Misuse Act (2011), any person who willfully, maliciously, and repeatedly uses electronic communication to harass another person and makes a threat with the intent to place that person in reasonable fear for his own safety or the safety of a member of that person's immediate family commits the crime of cyber stalking. On conviction, an offender is liable to a fine and/or imprisonment of up to 5 years.

In New Zealand, the Harmful Digital Communications Act (2015) makes it an offence to post a digital communication with the intent of harming the victim, or that causes harm to the victim and would do so to an "ordinary reasonable person" in that position.[938] The law encourages the court to consider the following factors when assessing whether a post would cause harm: the context of the communication, whether it was true or false, whether it was repeated, the extent of its circulation, the extremity of the content, anonymity, and the age and characteristics of the victim. The offence is punishable by up to a fine of NZ$50,000 (€27,961)[939] or up to 2 years imprisonment. The legislation also empowers the court to order the take-down of online content, the publication of a correction or apology, or a cease-and-desist order.

France has expanded the concept of online harassment causing harm through its laws on "moral harassment." The French Penal Code defines harassment as "repeated speech or behaviour" with the purpose or effect of causing a deterioration in the victim's way of life, resulting in an impairment of her rights and dignity or physical or mental health, or jeopardizing her career.[940] The offence is punishable by 1 year's imprisonment and a fine of €15,000. Based on 2014 and

[936] 1 NGN = 0.00234921 EUR on May 23, 2020.
[937] Nigeria, Cybercrime Act, 2015.
[938] New Zealand, Harmful Digital Communications Act, 2015. Available at: http://www.legislation.govt.nz/act/public/2015/0063/latest/whole.html#DLM5711852
[939] 1 NZD = 0.559223 EUR on May 23, 2020.
[940] Veronique Vincent, "France adopts a new law on sexual harassment," *Soulier Avocats*, Aug. 1, 2012. Available at: https://www.soulier-avocats.com/en/france-adopts-a-new-law-on-sexual-harassment/

2018 amendments to the law, the use of any "digital or electronic medium" to commit moral harassment is treated as an aggravating factor, resulting in a maximum penalty of 1 year's imprisonment and a fine of €30,000. Under the new version of the law, any form of harassment online, including harassment conducted through publicly accessible online platforms or direct private messaging, is treated as an aggravating factor in sentencing.[941]

2. Notable Cases

a. Norway: Norwegian Prosecution Authority v. X (2012)[942]

In this case, a Norwegian trial court sentenced a defendant to 30 days imprisonment for hate speech: an insulting post on the complainant's public Facebook profile.

The defendant posted a racist message on the public Facebook profile of a well-known singer and writer with Norwegian-African heritage, telling her to return to Africa, calling her the n-word, insulting her and comparing her to an animal. He was convicted of hate speech based under Section 185 of the Norwegian Penal Code.

The court emphasized that his actions were particularly harmful because the defendant posted on a profile set to public, thereby increasing the potential viewers of his post. The defendant received a sentence of thirty days unconditional imprisonment.

b. Singapore: Ye Lin Myint v. Public Prosecutor (2019)[943]

In this case, the Supreme Court of Singapore upheld a sentence of over 2 years in prison for defendant who waged an intimidation campaign that included threatening emails. Of the 2-year sentence, 6 months were for the offense of intentional harassment.

In 2017, a former insurance agent in Singapore waged a campaign of criminal intimidation against current and potential clients that failed to show up to appointments, by sending letters and

941 French Penal Code, Article 222-33-2-2-4 .
942 *Norwegian Prosecution Authority v. X*, 16-051378MED-OTIR/08 (Oslo Dist. Ct. 2016). Available at: https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2016/09/Norway-v-X.pdf
943 *Ye Lin Myint v Public Prosecutor*, SGHC 221 (Singapore Sup. Ct. 2019). Available at: https://www.supremecourt.gov.sg/docs/default-source/module-document/judgement/gd-for-ma-9029-of-2019-(final)-pdf.pdf

emails that threatened to humiliate the recipients, make them jobless, destroy their reputations, or even physically harm them or their family members. The defendant even began similarly harassing family members and neighbors of the first set of victims.

In finding sufficient grounds for conviction of intentional harassment (under Article 3(2) of the Prevention of Online Harassment Act) and criminal intimidation by anonymous communication (under Penal Code § 507), the judge took into account the following factors: (1) the alarm caused to the victims; (2) the significant public disquiet caused; (3) the appellant's serial offenses; and (4) the fact that the harassing communications had been made anonymously. The High Court ultimately approved an aggregate sentence of 29 months imprisonment for the above charges (6 months for intentionally causing harassment, alarm, or distress).

c. Uganda: Nyanzi v. Uganda (2019)[944]

In this case, a Ugandan trial court sentenced an academic to 15 months in prison for a Facebook post in which she insulted the Ugandan president. Notably, as discussed in more detail above in the "Criticizing Public Figures" section, a different case involving Nyanzi resulted in a finding by the OHCR's Working Group on Arbitrary Detention that her prosecution for cyberharassment violated free expression.[945]

Ugandan human rights activist and academic Stella Nyanzi was charged with cyberharassment for a Facebook post in which she blamed Ugandan President Yoweri Museveni for the corruption of Ugandan public institutions and expressed a wish that he had been burned up by the "acidic pus" in his mother's birth canal. The prosecution described Nyanzi's post as a "brutish attack on the person of the president and his late mother" involving "immoral words," consistent with the meaning of harassment.[946]

944 *See Dr. Stella Nyanzi v. Uganda*, UGHCCRD 39 (High Court of Uganda Hoi Den at Kampala 2019). Available at: https://ulii.org/ug/judgment/hc-criminal-division-uganda/2019/39

945 *Opinion No. 57/2017 concerning Stella Nyanzi* (Human Rights Council Working Group on Arbitrary Detention 2017). Available at: https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2017/04/A_HRC_WGAD_2017_57.pdf

946 "Ugandan academic Stella Nyanzi jailed for 'harassing' Museveni," *Al-Jazeera*, Aug. 3, 2019. Available at: https://www.aljazeera.com/news/2019/08/ugandan-academic-stella-nyanzi-jailed-harassing-museveni-190803141817222.html

Under the Ugandan Computer Misuse Act of 2011, any use of a computer to "mak[e] any request, suggestion or proposal which is obscene, lewd, lascivious or indecent" can be deemed cyberharassment with a sentence not exceeding three years. Accordingly, Nyanzi was convicted of cyberharassment and sentenced to 18 months in prison.

3. Exceptional Cases

In the two cases summarized below, national courts found criminal prosecutions under harassment laws to infringe upon free expression as defined by national constitutional law. These cases depart from the general absence of tensions observed in court cases between the right to freedom of expression and prosecutions for cyberharassment.

a. Canada: R v. Gregory Alan Elliott (2016)[947]

A Canadian trial court dismissed criminal harassment charges because the defendant's homophobic Tweets did not establish complainant's reasonable fear for their safety. The court referenced the Canadian right of freedom of expression as including the right to express offensive opinions.

In 2012, Canadian citizen Gregory Alan Elliott tweeted offensive and homophobic messages targeting two women's rights activists (Stephanie Guthrie and Heather Reilly). They blocked him after multiple exchanges, but he continued to tweet about them and at them through hashtags they frequently used, despite them asking him to stop.

The Ontario Court of Justice dismissed the activists' criminal harassment charges because the activists did not establish a reasonable fear for their safety. The Crown has to prove five elements in order to establish a criminal harassment charge under Section 264 of the Canadian Criminal Code: (i) repeated communication, (ii) harassment of the complainant, (iii) the defendant's awareness of said harassment, (iv) that the communication caused complainant(s) to fear for their safety, and (v) that this fear was reasonable.[948]

947 *R. v. Elliott*, 2016 ONCJ 35 (Ontario Ct. of Justice, 2016). Available at: https://www.canlii.org/en/on/oncj/doc/2016/2016oncj35/2016oncj35.html

948 Canadian Criminal Code, § 264. Available at: https://www.canlii.org/en/ca/laws/stat/rsc-1985-c-c-46/latest/rsc-1985-c-c-46.html#sec264_smooth

The court concluded that Elliot's actions towards both women satisfied the first two criteria: there was repeated communication in the form of harassment, as the communication was directed at the two women, even though it took place via public hashtags. The Judge found that the third criteria of the defendant's awareness of the harassment was met only in Reilly's case, as she told Elliott to stop contacting her, while Guthrie did not. Guthrie only blocked him without telling him that she had blocked him; therefore, the judge found "no direct evidence of Mr. Elliott's knowledge of Ms. Guthrie's harassment.

The court found the fourth and fifth elements of the offense lacking. The Judge was not satisfied that Reilly feared for her safety. In Guthrie's case, although the Judge agreed she had feared for her safety, he found this fear not reasonable, writing: "[h]ad there been anything in the tweets of a violent or sexual nature or that indicated the irrationality that Ms. Guthrie perceived, that could support a fear of danger on the basis that he would be capable of anything," but that no such tweets were in the record.

The Judge made two important points with regard to the freedom of expression. First, he considered the use of hashtags as "similar to announcing a public meeting," reasoning that "once someone creates a hashtag, anyone can use it," and unless everyone is able to use it freely, that will "limit the operation of Twitter in a way that is not consistent with freedom of expression." Second, in writing about why Elliott could not have been reasonably expected to know that he was harassing Guthrie, he stated that the Canadian tradition of freedom expression allowed for the defendant to hold "a controversial or even offensive opinion . . . use extreme, hyperbolic, provocative language such as 'fascist feminists,'. . . and be . . . homophobic and insulting." The judge reasoned that since the defendant's view as demonstrated by his Tweets was "that he could write what he wanted," and this view " conforms to the Twitter rules and the Canadian value of freedom of expression," the defendant "would not know that Ms. Guthrie was harassed by his doing what was lawful and what the platform they were both using permitted."

c. India: Singhal v. Union of India (2015)

In this case, the Indian Supreme Court struck down a provision criminalizing the electronic sending of "grossly offensive" messages under the Indian constitutional right to free expression.

The law failed for vagueness, but also failed to show a clear enough connection to public order; it did not distinguish between incitement leading to public disorder and mere advocacy of offensive points of view.[949]

Two woman posted Facebook comments criticizing the shutdown of Mumbai after a political leader died. Mumbai authorities arrested and charged them under Section 66A of the Information Technology Act of 2000 (ITA). This law establishes a criminal offense for anyone who electronically sends information (a) that is "grossly offensive," or (b) with knowledge of its falsity and for the purpose of causing "annoyance, inconvenience, danger, obstruction, insult, injury, hatred," etc. The prosecution against the women was dismissed, but the women filed a petition to the Supreme Court of India, challenging Section 661's validity on grounds of free expression.[950]

The Supreme Court ruled that Section 66A is unconstitutional in its entirety. First, the law fails to establish a clear proximate relation to the protection of public order. The section fails to distinguish between "mere discussion or advocacy of a particular point of view, which may be annoying or inconvenient or grossly offensive to some" and "incitement by which such words lead to an imminent causal connection with public disorder." Second, the section fails for vagueness; given the undefined and broad terms in the provision, accused people were not put on notice as to what speech would be legal versus illegal.

F. "Offline" Laws Applicable Online

Not every country has legislation that directly addresses cyberharassment. Where this is the case, however, the perpetrator may instead be committing a criminal offence under broader acts covering harassment, stalking, and so on, which can be applied offline and online.

949 *Shreya Singhal v. U.O.I*, Writ Petition No. 167 of 2012 (Sup. Ct. of India 2015). Available at: https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2015/06/Shreya_Singhal_vs_U.O.I_on_24_March_2015.pdf

950 "Shreya Singhal v. Union of India," Columbia Global Freedom of Expression.

1. National Laws

a. Australia

Various offences in Section 474 of the Australian Criminal Code could apply to cyberbullying or cyber abuse. These include using a "carriage service" to make a threat (Section 474.15), hoax threat (Section 474.16), to "menace, harass or cause offence" (Section 474.17) or "for suicide related material" (Section 474.15). Additionally, Section 272.14 makes "using a telecommunications network with intent to commit a serious offence" a crime as well.[951]

The Attorney-General's Department has noted that Section 474.17's provision on using a carriage service to "meance, harass, or cause offence," which can result in up to 3 years of imprisonment, does not define these terms, thereby requiring the application of "community standards and common sense" to understand their meaning. Along those lines, the Department has stressed that it is important to consider and balance various factors when applying this law, including "standards of morality, decency, and propriety generally accepted by reasonable adults," whether the material has "literary, artistic, or educational merit," and the material's "general character . . . including whether it is of a medical, legal, or scientific character."

Additionally, that different territories and states in Australia have different laws that could apply to cyberharassment. For instance, Brodie's Law in Victoria criminalizes cyberbullying, making it an offence punishable by up to ten years imprisonment, via extending the application of stalking provisions in Victoria's Crimes Act (1958).[952]

b. Canada

Besides the revenge porn law discussed above, Canada does not have any other laws that directly mention cyberharassment. That being said, there are many potentially applicable provisions in the Criminal Code.[953] These include criminal harassment (Section 264), uttering

[951] Australian Criminal Code, Chapter 3, Criminal offences for cyberbullying. Available at: https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Legal_and_Constitutional_Affairs/Cyberbullying/Report/c03

[952] "Bullying - Brodie's Law," Victoria State Government: Justice and Community Safety, 2011. Available at: https://www.justice.vic.gov.au/safer-communities/crime-prevention/bullying-brodies-law

[953] "Cyberbullying and the Non-consensual Distribution of Intimate Images," Department of Justice, Government of Canada, last modified 2017. Available at: https://www.justice.gc.ca/eng/rp-pr/other-autre/cndii-cdncii/p4.html

threats (Section 264.1), intimidation (Section 423(1)), mischief in relation to data (Section 372), false messages and "indecent or harassing telephone calls" (Section 372), incitement of hatred (Section 319), and defamatory libel (Section 289–301).

c. United Kingdom

Similar to Canada, the United Kingdom does not have laws that specifically target cyberharassment other than revenge porn. However, the Crown Prosecution Service has provided guidelines on prosecuting cases involving social media communications which point to other criminal offenses that may be used to prosecute actions that may constitute cyberbullying.[954] Examples include threatening to kill someone, threatening to commit criminal damage, and publishing material that may lead to the identification of the victim of a sexual offense.[955]

G. Social Media Guidelines

This survey of social media company guidelines draws from the content policies of a wide range of online speech platforms. These include Facebook, Google, Instagram, Reddit, Snapchat, Twitter, Tumblr, WhatsApp, and YouTube. The below observations should not be seen as exhaustive, as platform guidelines diverge considerably in terms of thoroughness, activities covered, and definitions. Notably, there is no clear overarching definition of hateful conduct in social media company guidelines. Furthermore, there are many overlaps between hateful conduct and other areas of concern online, such as terrorist activities.

Broadly speaking, these platforms prohibit users from engaging in threats of violence or property damage, cyberharassment or cyberbullying, doxing of private personally identifiable information, and hate speech. Additionally, most platforms—similar to most national and supranational courts—consider context (e.g. purpose, audience, speaker) when interpreting and applying these guidelines.

954 "Social Media - Guidelines on prosecuting cases involving communications sent via social media," The Crown Prosecution Service, Aug. 21, 2018. https://www.cps.gov.uk/legal-guidance/social-media-guidelines-prosecuting-cases-involving-communications-sent-social-media

955 Offences Against the Person Act (1861), § 16; Criminal Damage Act (1971), § 25; Sexual Offences (Amendment) Act (1992), § 5.

First, platforms prohibit users from posting threats, whether against individuals or groups of people. This often extends beyond prohibiting threats of physical violence to prohibiting threats of property damage, theft, or financial harm. Some platform policies also mention sexual threats.

Second, policies related to cyberharassment ban content that harasses or bullies others. Platforms also generally prohibit doxing, forbidding users from revealing private information of others, including medical records, financial information, social security numbers, and passport numbers, as well as unlisted contact information like email or physical addresses and phone numbers.

Third, platform policies relating to hate speech prohibit directed attacks based on protected characteristics. Platforms define this category in different ways. Twitter defines hateful conduct as promoting violence against or "directly attack[ing] or threaten[ing]" others on the basis of protected characteristics. Facebook defines an attack as "violent or dehumanizing speech, statements of inferiority, or calls for exclusion or segregation." Snapchat describes this category as "content that demeans, defames, or promotes discrimination or violence" on the basis of a protected characteristic.

Notably, social media companies have embraced a much *broader* understanding of protected groups than those identified in international human rights treaties. Taken together, the ICCPR, ICERD, and ACHR encourage prohibiting the advocacy of hatred of a national, racial, or religious nature that constitutes incitement (e.g. race, color, ethnicity, religion, national origin). In contrast, social media companies incorporated these understandings of protected groups or factors, but also included sexual orientation, caste, gender, disability, immigration status, and veteran status.

Some academics have advocated for an even more flexible approach. A book by Teo Keipi and his colleagues on online hate and harmful content stresses the fluidity of characteristics that might make a group the target of hate, and broaden the range of actions considered to constitute targeting. The authors state that "online hate expression" should be understood as the use of information and communication technology to "advocate violence against, separation from,

defamation of, deception about, or hostility towards others."[956] They emphasize that although these attacks are usually based on protected defining characteristics of a group, such as race or sexual orientation, they can also be motivated by global developments related to identity. For instance, a refugee crisis may make refugee status a defining characteristic of a group.

Activities That May Constitute Cyberharassment

1. Threats
    a. Twitter: "You may not threaten violence against an individual or a group of people . . . We define violent threats as statements of an intent to kill or inflict serious physical harm on a specific person or group of people."[957]
    b. Tumblr: "Don't post content which includes violent threats towards individuals or groups – this includes threats of theft, property damage, or financial harm… Don't post content that encourages or incites violence…"[958]
    c. Instagram: "Serious threats of harm to public and personal safety aren't allowed. This includes specific threats of physical harm as well as threats of theft, vandalism and other financial harm."[959]
    d. YouTube: "Content that threatens individuals is not allowed on YouTube."[960]

2. Cyberharassment
    a. YouTube: "We . . . do not allow content that targets an individual with prolonged or malicious insults based on intrinsic attributes, including their protected group status or physical traits."[961]

[956] Teo Keipi et al., *Online Hate and Harmful Content: Cross-National Perspectives*, 1 Edition, Routledge Advances in Sociology (London : New York: Routledge, Taylor & Francis Group, 2017).
[957] "The Twitter Rules," Twitter, 2020, https://help.twitter.com/en/rules-and-policies/twitter-rules.
[958] "Community Guidelines," Tumblr, Jan. 23, 2020, https://www.tumblr.com/policy/en/community.
[959] "Community Guidelines," Instagram, 2020, https://help.instagram.com/477434105621119.
[960] "Policies and Safety," YouTube, 2020, https://www.youtube.com/about/policies/#community-guidelines.
[961] "Policies and Safety," YouTube, 2020. Available at: https://www.youtube.com/about/policies/#community-guidelines

b. Reddit: "We do not tolerate the harassment, threatening, or bullying of people on our site; nor do we tolerate communities dedicated to this behavior."[962]
c. Snapchat: "We have zero tolerance for bullying or harassment of any kind."[963]
d. Google: "We do not allow content that sends messages intended to harass, bully, or physically or sexually threaten others."[964]
e. TikTok: "Users should feel safe to express themselves without fear of being shamed, humiliated, bullied, or harassed."[965]
f. WhatsApp: "You will not use (or assist others in using) our Services in ways that: . . . (b) are illegal, obscene, defamatory, threatening, intimidating, harassing, hateful, racially, or ethnically offensive, or instigate or encourage conduct that would be illegal, or otherwise inappropriate, including promoting violent crimes."[966]
g. Twitter: "While some consensual nudity and adult content is permitted on Twitter, we prohibit unwanted sexual advances and content that sexually objectifies an individual without their consent. This includes, but is not limited to:
    i. sending someone unsolicited and/or unwanted adult media, including images, videos, and GIFs;
    ii. unwanted sexual discussion of someone's body;
    iii. solicitation of sexual acts; and
    iv. any other content that otherwise sexualizes an individual without their consent." [967]

3. Releasing private (especially identifying) information about someone

[962] "Reddit Content Policy," Reddit, 2020. Available at: https://www.redditinc.com/policies/content-policy
[963] "Community Guidelines," Snap Inc., 2019. Available at: https://www.snap.com/en-US/community-guidelines
[964] "Community Guidelines," Google.
[965] "Community Guidelines," TikTok.
[966] "WhatsApp Legal Info," WhatsApp, 2019. Available at: https://www.whatsapp.com/legal/
[967] "The Twitter Rules," Twitter.

a. Tumblr: “Don't post content that violates anyone's privacy, especially personally identifying or confidential information like credit card numbers, social security numbers, or unlisted contact information.”[968]
b. YouTube: (as an example of a violation of the harassment and cyberbullying policy) “Revealing someone’s private information, such as their home address, email addresses, sign-in credentials, phone numbers, passport number, or bank account information.”[969]
c. Google: “We do not allow the sharing of a private person's confidential and personally identifiable information (e.g. medical records or financial information).”[970]

4. Hate speech (i.e. a directed attack based on protected characteristics)
    a. Facebook: “We do not allow hate speech . . . [defined] as a direct attack on people based on what we call protected characteristics — race, ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity, and serious disease or disability. We also provide some protections for immigration status. We define attack as violent or dehumanizing speech, statements of inferiority, or calls for exclusion or segregation.”[971]
    b. Snapchat: “Don't post any content that demeans, defames, or promotes discrimination or violence on the basis of race, ethnicity, national origin, religion, sexual orientation, gender identity, disability, or veteran status.”[972]
    c. Twitter: “Hateful conduct: You may not promote violence against or directly attack or threaten other people on the basis of race, ethnicity, national origin, caste, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious

968 “Community Guidelines,” Tumblr.
969 “Policies and Safety,” YouTube.
970 “Community Guidelines,” Google.
971 “Community Standards,” Facebook.
972 “Community Guidelines,” Snap Inc.

disease. We also do not allow accounts whose primary purpose is inciting harm towards others on the basis of these categories."[973]

H. Conclusion

There are four main insights to be drawn from legal developments regarding cyberharassment.

First, there is no clear definition of cyberharassment and, except for a few mentions of harassment in the workplace or on the basis of protected characteristics such as gender, the subject is not directly addressed in international or regional human rights treaties.

Second, this research has identified four main strands of cyberharassment: cyberstalking, sexual harassment (often via revenge porn), cyberbullying involving minors, and other cyber activity that causes or is likely to cause harm. Other than cyberbullying involving minors, these activities tend to be criminalized, though to varying degrees. Online harassment involving minors is generally not criminalized *in and of itself* with the exception of a number of US states. Most countries have concluded that the education system is a better avenue to deal with the problem of cyberbullying than the criminal justice system.

Third, since cyberharassment is a relatively recent phenomenon, many countries have yet to create laws that specifically address the issue. However, many of them, including Canada, Australia, and the United Kingdom, have "offline" laws, such as those against threats, which can be applied to prosecute cyberharassment.

Fourth, there are very few instances of criminal responsibility for cyberharassment being held to violate freedom of expression. Our research has unearthed only three such cases: one decided under international principles by the OHCR's Working Group on Arbitrary Detention, in *Nyanzi v. Uganda*,[974] and two decided by national courts under constitutional principles, the *Elliott*

[973] "The Twitter Rules," Twitter.
[974] *Opinion No. 57/2017 concerning Stella Nyanzi*, A/HRC/WGAD/2017/57 (Human Rights Council Working Group on Arbitrary Detention 2017).

case in Canada and *Singhal* case in India. In all three, the speech for which defendants were prosecuted was politically charged.

This absence of court cases finding that cyberharassment prohibitions violate free expression may be because cyberharassment is generally a more straightforward personal crime than in the politics-charged *Elliot* and *Singhal* cases. In most cases, cyberharassment offenses center on the relationship between two individuals rather than broader social or political commentary.

Alternatively, the lack of court cases finding criminalization of cyberharassment to violate free expression may be due to the novel nature of cyberharassment. As this field of law develops further, international jurisprudence on freedom of expression may generate more constraints on criminalizing cyberharassment.