

**THE CONSEQUENCES OF INACTION: AN
INQUIRY INTO INTERNATIONAL CRIMINAL
LIABILITY OF SOCIAL MEDIA COMPANIES
FOR ARTSAKH 2020¹**
Rajika L. Shah²

TABLE OF CONTENTS

- I. INTRODUCTION
- II. BACKGROUND: A HISTORY OF TENSION
- III. SOCIAL MEDIA AND THE 2020 CONFLICT
 - A. MANIPULATED SOCIAL MEDIA PROMOTED VIOLENT ANTI-ARMENIAN RHETORIC
 - B. THE RESPONSE FROM SOCIAL MEDIA COMPANIES WAS SLOW AND INEFFECTIVE
- IV. INTERNATIONAL CRIMINAL TRIBUNALS HAVE HELD COMPANIES AND MASS MEDIA EXECUTIVES LIABLE FOR THEIR ROLE IN ATROCITY CRIMES
 - A. THE NUREMBERG IMT FOUND LIABILITY WHERE A NEWSPAPER PUBLISHER CONTINUED TO PUBLISH ARTICLES INCITING GENOCIDAL VIOLENCE WHILE AWARE OF THE STATE’S GENOCIDAL VIOLENCE
 - B. THE SUBSEQUENT NUREMBERG TRIALS FOUND SECONDARY CORPORATE LIABILITY WHERE A COMPANY KNOWINGLY SUPPLIED A COMMODITY TO THE STATE WHILE AWARE THE STATE WAS USING THE COMMODITY TO COMMIT SERIOUS CRIMES
 - C. THE ICTR’S MEDIA CASE HELD MEDIA COMPANIES EXECUTIVES LIABLE FOR INCITING GENOCIDE DUE TO THE MESSAGING DISSEMINATED ON THEIR PLATFORMS
 - 1. *RTLM Radio Broadcasts Advocated for the Extermination of Tutsis*
 - 2. *Kangura Newspaper Articles Similarly Demonized Tutsis*
 - 3. *The ICTR Trial Chamber Found Individual Criminal Responsibility for Media Company Heads Who Fanned the Flames of Violence*

¹ This report was prepared at the request of the Armenian Bar Association. It was researched and drafted in the spring semester of 2021 by Meighann Mahoney, Sarina Piliguin, and Tamar Thomassian, students in the Loyola Genocide Justice Clinic at LMU Loyola Law School, which was at that time a program of the Loyola Center for the Study of Law and Genocide (“CSLG”). CSLG 2021 summer interns Rebecca Brown, Garen Kosoyan, and Angelina Sophonpanich provided additional research assistance. The author presented the report’s findings at the conference “In the Shadow of Territorial Conflict: Artsakh and Other Legacies of Soviet-era Media Control and Speech Norms,” held at Southwestern Law School on February 4, 2023, and finalized the report for publication.

² Visiting Associate Clinical Professor and Director, Loyola Justice for Atrocities Clinic (formerly the Loyola Genocide Justice Clinic), LMU Loyola Law School.

4. *The ICTR Appeals Chamber Affirmed Media Executives' Responsibility to Prevent the Spread of Violent Content*
- V. THE INTERNATIONAL COMMUNITY RECOGNIZED FACEBOOK'S ROLE IN EXACERBATING ETHNIC TENSIONS IN MYANMAR
 - A. FACEBOOK PLAYED A CRUCIAL ROLE IN FACILITATING VIOLENCE AGAINST THE ROHINGYA
 - B. FACEBOOK'S EFFORTS TO CURB USE OF ITS PLATFORM TO SPREAD VIOLENCE WERE SEVERELY LACKING
 - C. A HUMAN RIGHTS IMPACT ASSESSMENT REPORT COMMISSIONED BY FACEBOOK DOWNPLAYED ITS RESPONSIBILITY FOR THE VIOLENCE IN MYANMAR
 - D. EVEN AFTER IT WAS ON NOTICE, FACEBOOK CONTINUED TO BE USED FOR HATE SPEECH IN MYANMAR
- VI. CONCLUSION

I. INTRODUCTION

On July 12, 2020, clashes between Armenian and Azerbaijani forces began along their respective northern borders.³ The several days of fighting proved to be the most serious and deadliest escalation of hostilities between the two nations since the Four-Day War in April 2016.⁴ In Azerbaijan, tens of thousands of people gathered in the capital Baku, demanding the government declare war against Armenia.⁵

Two months later, on September 27, 2020, Azerbaijan, with the help of Turkish military aid and equipment, initiated a military attack in the territory of Artsakh, or Nagorno-Karabakh,⁶ a geographically isolated enclave with a predominantly ethnic Armenian population that falls within Azerbaijan's borders, but has maintained a de facto separatist government known as the Republic of Artsakh since the

³ Associated Press, *Clashes Resume on Armenian-Azerbaijani Border*, WASH. POST (July 16, 2020, 6:38 AM EDT), https://www.washingtonpost.com/world/europe/clashes-resume-on-armenian-azerbaijani-border/2020/07/16/7c858fa6-c750-11ea-a825-8722004e4150_story.html.

⁴ *Id.*

⁵ *Azerbaijan Protestors Demand War After Armenia Clashes*, BBC NEWS (July 15, 2020), <https://www.bbc.com/news/world-europe-53415693>.

⁶ This report may refer interchangeably to both terms to describe the same region.

early 1990s.⁷ Shortly thereafter, both the Republic of Artsakh and Armenia issued statements announcing the general mobilization of troops in defense.⁸ Azerbaijan reportedly used missiles, aerial drones, cluster munitions, and phosphorus bombs in attacks on Artsakh.⁹ Many attacks strategically targeted civilians and key medical and civilian infrastructure such as hospitals, churches, and schools.¹⁰ Azerbaijani forces also launched direct attacks on Armenia, targeting both military and civilian infrastructure.¹¹ Active fighting ended when the parties signed a Russia-brokered peace treaty on November 10, 2020.¹² For purposes of this Report, the entire forty-four-day war is called the “2020 Conflict.”

Before, during, and after the 2020 Conflict, accounts on social media platforms, including Instagram, TikTok, Facebook, Twitter, and Reddit, were used as a primary means of disseminating conflict-related information—including misinformation, disinformation, and hate

⁷ As a result of a subsequent Azerbaijani offensive in Nagorno-Karabakh that began on September 19, 2023, virtually all of the indigenous ethnic Armenians in the territory fled to Armenia out of fear for their safety and what the future might hold under Azerbaijani government rule. See Joel Gunter, *Deserted Nagorno-Karabakh Reveals Aftermath of Lightning-Fast Armenian Defeat*, BBC NEWS (Oct. 3, 2023), <https://bbc.com/news/world-europe-66995976>. The government of the Republic of Artsakh ceased to exist as of January 1, 2024. George Wright, *Nagorno-Karabakh: Armenia Says 100,000 Refugees Flee Region*, BBC NEWS (Sept. 30, 2023), <https://www.bbc.com/news/world-europe-66969845>.

⁸ *Azerbaijan Launches Wide Scale Offensive*, EVN REP. (Sept. 27, 2020), <https://www.evnreport.com/spotlight-karabakh/azerbaijan-launches-wide-scale-offensive>.

⁹ *Azerbaijan: Unlawful Strikes in Nagorno-Karabakh*, HUM. RTS. WATCH (Dec. 11, 2020, 3:00 AM), <https://www.hrw.org/news/2020/12/11/azerbaijan-unlawful-strikes-nagorno-karabakh>.

¹⁰ *Id.*

¹¹ Hugh Williamson & Tanya Lokshina, *Unlawful Attacks on Medical Facilities and Personnel in Nagorno-Karabakh*, HUM. RTS. WATCH (Feb. 26, 2021), <https://www.hrw.org/news/2021/02/26/unlawful-attacks-medical-facilities-and-personnel-nagorno-karabakh>.

¹² Robyn Dixon, *Cease-Fire in Nagorno-Karabakh Provokes Protests in Armenia, Celebrations in Azerbaijan*, WASH. POST (Nov. 10, 2020, 2:51 PM EST), https://www.washingtonpost.com/world/europe/nagorno-karabakh-ceasefire-armenia-russia-azerbaijan/2020/11/10/b1b9bcc0-231b-11eb-9c4a-0dc6242c4814_story.html.

propaganda¹³—throughout the region and worldwide.¹⁴ Though some of the posts shared information such as status updates and the current situation of the civilian population, widely disseminated social media postings also included a range of disinformation that severely inflamed emotions and contributed to further violence in the region.

Several posts included hate speech referring to Armenians as “dogs,” “wild beasts,” and “rats” who should be driven out of Nagorno-Karabakh, which has an indigenous ethnic Armenian population.¹⁵ Disturbingly, extremely graphic images and videos of Azerbaijani soldiers appearing to commit horrific war crimes against ethnic Armenians also began spreading on social media.¹⁶ This was

¹³ For the purposes of this Report, disinformation is “verifiably false or misleading information . . . created, presented and disseminated . . . to *intentionally* deceive the public” and “[m]ay cause public harm,” while misinformation may be accidental. See European Commission, *Code of Practice on Disinformation*, EUR. UNION (Sept. 2018), <https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation> (emphasis added). Propaganda is a broader concept that “can be described as a method of communication, by State organs or individuals, aimed at influencing and manipulating the behaviour of people in a certain predefined way” and thus contains a manipulative aspect. Eric De Brabandere, *Propaganda*, in MAX PLANCK ENCYCLOPEDIA OF PUBLIC INTERNATIONAL LAW (Rüdiger Wolfrum ed., 2019), ¶ 1. “Discriminatory propaganda” or “hate propaganda” incites atrocities and is related to hate speech. *Id.* at ¶ 21. The United Nations Office on Genocide Prevention and the Responsibility to Protect defines hate speech as “any kind of communication in speech, writing or behaviour, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of . . . religion, ethnicity, nationality, race, colour, descent, gender or other identity factor.” Office on Genocide Prevention and the Responsibility to Protect, *Strategy and Plan of Action on Hate Speech*, U.N. (May 2019), <https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20Plan%20of%20Action%20on%20Hate%20Speech%2018%20June%20SYNOPSIS.pdf>.

¹⁴ Katy Pearce, *While Armenia and Azerbaijan fought over Nagorno-Karabakh, their citizens battled on social media*, WASH. POST (Dec. 4, 2020, 7:45 AM EST), <https://www.washingtonpost.com/politics/2020/12/04/while-armenia-azerbaijan-fought-over-nagorno-karabakh-their-citizens-battled-social-media/>.

¹⁵ See Joe Nersessian, *The Mixed Messaging of Ilham Aliyev*, EVN REP. (Oct. 22, 2020), <https://www.evnreport.com/politics/the-mixed-messaging-of-ilham-aliyev> (quoting English translations of numerous speeches given by Azerbaijani President Ilham Aliyev before and during the 2020 Conflict).

¹⁶ See Andrew Roth, *Two Men Beheaded in Videos from Nagorno-Karabakh War Identified*, GUARDIAN (Dec. 15, 2020), <https://www.theguardian.com/world/2020/dec/15/two-men-beheaded-in-videos-from-nagorno-karabakh-war-identified>; see also Ulkar Natiqqizi & Joshua Kucera,

not surprising given the long-standing policy and practice by government authorities in Azerbaijan to “deliberately amplif[y] and exacerbate[.]” the effects of past conflict dating back to the post-Soviet era by actively stoking anti-Armenian hatred and fear among the people of Azerbaijan and repressing freedom of the press.¹⁷

Social media posts stirring up nationalist sentiment simplified the narrative and “contributed to the deepening of hatreds and dehumanization of the other.”¹⁸ This manifested in real life as hate crimes against Armenian communities around the world began to increase. For example, (i) on September 17, 2020, the Armenian Church of St. Gregory in San Francisco, California, was burned in a suspected case of arson;¹⁹ (ii) on October 28, 2020, a group of Turkish nationals known as the Grey-Wolves took to the streets of Lyon, France looking for Armenians to kill;²⁰ and (iii) on October 29, 2020,

Evidence of Widespread Atrocities Emerges Following Karabakh War, EURASIANET (Dec. 9, 2020), <https://eurasianet.org/evidence-of-widespread-atrocities-emerges-following-karabakh-war>.

¹⁷ Roza Malkumyan, *Baku’s Hostility Has Not Abated since the Fall of Nagorno-Karabakh*, FREEDOM HOUSE (Nov. 30, 2023), <https://freedomhouse.org/article/bakus-hostility-has-not-abated-fall-nagorno-karabakh>. In 2020, Azerbaijan ranked 168th out of 180 countries on Reporters Without Borders (RSF) World Press Freedom Index, while Armenia ranked 61st. *World Press Freedom Index*, RSF, <https://rsf.org/en> (last visited June 12, 2024).

¹⁸ See EUR. RES. FOR MEDIATION SUPPORT, MEDIA AND DISINFORMATION IN THE NAGORNO-KARABAKH CONFLICT AND THEIR ROLE IN CONFLICT RESOLUTION AND PEACEBUILDING, (Jan. 2021) (report on exploratory seminar held on Dec. 17, 2020), at 9-10 (hereinafter, “ERMES Report”).

¹⁹ *Fire Burns Armenian Church Building Overnight in San Francisco; Arson Suspected*, CBS S.F. (Sept. 17, 2020), <https://sanfrancisco.cbslocal.com/2020/09/17/armenian-church-burns-san-francisco-arson-suspected/>.

²⁰ Tim Hume, *Turkish Ultranationalist Group Linked to “Hunt For Armenians” in France*, VICE (Oct. 29, 2020, 3:47 PM), <https://www.vice.com/en/article/epddna/turkey-france-armenia-grey-wolves-lyon>; see also Patrick Keddie, *France has Banned the ‘Grey Wolves’ – But Who are They?*, AL JAZEERA (Nov. 24, 2020), <https://www.aljazeera.com/features/2020/11/24/france-has-banned-the-grey-wolves-but-who-are-they>.

three Armenian men were stabbed in Fresno, California, at a “Defend Armenia” rally.²¹

Despite the large number of these types of posts and their widespread circulation in the context of a long-simmering conflict, major social media companies—all of whom had policies in place in 2020 regarding the dissemination of hate speech and promotion of violence—appeared unprepared to handle the “rapid-fire dissemination” of “nationalist and ultra-nationalist narratives . . . across social media [that] often resemble[d] those from violent extremist groups.”²²

This Report analyzes the potential liability of social media companies—whose ubiquitousness and influence today as the primary means of communication for billions of smartphone users around the world are unrivaled by any other form of mass media—and/or the decisionmakers at such companies under international criminal law for the consequences of failing to prevent the spread of disinformation and hate speech on their platforms during the 2020 Conflict.²³

Part II of this Report provides a brief background to the long history of tension in the region, particularly as understood in the context of the early twentieth century Armenian Genocide. Part III reviews the role of inflammatory social media posts before, during, and after the 2020 Conflict. Part IV examines the international criminal precedent for mass media actors and companies who play a role in inciting atrocities. Part V considers the scrutiny Facebook received from the UN-authorized Independent International Fact-Finding Mission on Myanmar for the role its platform played in the extreme violence targeting Rohingya Muslims in 2017. Finally, Part VI concludes that, while top social media executives cannot be held criminally liable in international law for the specific anti-Armenian hate speech posted on their platforms in relation to the 2020 Conflict because they did not personally endorse and amplify such speech, social media companies certainly could have done more to implement measures they knew or should have known by at least 2018 would help

²¹ Sara Sandrik, *Defend Armenia Protesters Call Fresno Stabbing Attack a Hate Crime*, ABC 30 (Oct. 29, 2020), <https://abc30.com/defend-armenia-fresno-rally-river-park-stabbing/7463718/>.

²² See ERMES Report, *supra* note 18, at 10.

²³ This report does not address any potential civil liability that may arise for social media companies or their key decision-makers in connection with the 2020 Conflict, whether in the United States or internationally.

stem the outpouring of violent content during conflict situations. Their failure to do so in 2020 resulted in exacerbated tensions and nurtured a breeding ground for atrocity crimes to occur.

II. BACKGROUND: A HISTORY OF TENSION

There is a long history of tension in the South Caucasus regarding Nagorno-Karabakh. Though the mountainous region was populated for centuries by both Christian Armenians and Turkic Muslim Azeris, by the late nineteenth century, when it was overtaken by the Russian empire, Nagorno-Karabakh had a majority ethnic Armenian population.²⁴ After the Bolshevik revolution in the early 1920s, Nagorno-Karabakh was established as an autonomous region of the Soviet Union, but within the borders of Soviet Azerbaijan.²⁵

As the Soviet Union moved toward collapse in the late 1980s, a separatist movement developed, and fighting broke out in Nagorno-Karabakh in 1988.²⁶ Although Nagorno-Karabakh petitioned to become part of Armenia when the Soviet Union fell, it ultimately remained within Azerbaijan's borders.²⁷ In 1991, the separatists declared themselves the independent Republic of Artsakh.²⁸ By the time a cease-fire took hold in early 1994, "separatists, with Armenian support, controlled Nagorno-Karabakh and seven surrounding Azerbaijani territories, constituting a total of 14 percent of Azerbaijan's overall territory."²⁹ Known as the first Karabakh war, it left between 15,000-30,000 people dead and more than a million displaced.³⁰ Though a formal mediation and peace process was

²⁴ *Nagorno-Karabakh Profile*, BBC (Jan. 30, 2024), <https://www.bbc.com/news/world-europe-18270325>; see also *Armenia*, CIA: THE WORLD FACTBOOK, <https://www.cia.gov/the-world-factbook/countries/armenia> (last updated Oct. 1, 2024).

²⁵ *Nagorno-Karabakh Profile*, *supra* note 24; see also CIA: THE WORLD FACTBOOK, *supra* note 24.

²⁶ CIA: THE WORLD FACTBOOK, *supra* note 24.

²⁷ See PATRICIA CARLEY, U.S. INST. OF PEACE, *NAGORNO-KARABAKH: SEARCHING FOR A SOLUTION* (1998).

²⁸ *Nagorno-Karabakh Profile*, *supra* note 24.

²⁹ CIA: THE WORLD FACTBOOK, *supra* note 24.

³⁰ See CARLEY, *supra* note 27; Mathieu Droin et al., *A Renewed Nagorno-Karabakh Conflict: Reading Between the Front Lines*, CTR. FOR STRATEGIC &

established in 1992 through the Minsk Group of the Organization for Security and Cooperation in Europe (OSCE), it was largely ineffective, and a proposed settlement plan collapsed in early 1998.³¹ The situation remained in a “simmering stalemate . . . punctuated by armed clashes” until 2020.³²

These tensions occurred within the larger backdrop of the Armenian Genocide carried out by Ottoman Turks from 1915 to 1923. During that genocide, as many as 1.5 million ethnic Armenians living in the Ottoman Empire were murdered and expelled from their homes, forced to march hundreds of miles with little to no food, water, or shelter from Eastern Anatolia into the Mesopotamian desert.³³ Those most responsible for the Armenian Genocide were never held legally accountable, and modern Turkey (the successor state to the Ottoman Empire) has consistently refused to acknowledge the massacres of the Armenians as a genocide.³⁴ Since Azeris are Turkic Muslims with close ties to Turkey, and Armenia is geographically bound by Turkey to the west and Azerbaijan to the east, there is a strong sense that the past, with all its hatreds and suspicion, is still very much alive.³⁵

III. SOCIAL MEDIA AND THE 2020 CONFLICT

A. MANIPULATED SOCIAL MEDIA PROMOTED VIOLENT ANTI-ARMENIAN RHETORIC

Before, during, and after the 2020 Conflict, numerous observers documented a variety of tactics utilized by Azerbaijan and Turkey to inflame anti-Armenian sentiment and shape public opinion in favor of the conflict.

As early as 2012, research showed that the online discourse in Azerbaijan consisted largely of “hate blogs” expressing hatred and propaganda against Armenians while promoting positions that were

INT'L STUD. (Sept. 22, 2023), <https://www.csis.org/analysis/renewed-nagorno-karabakh-conflict-reading-between-front-lines>.

³¹ See CARLEY, *supra* note 27.

³² *Nagorno-Karabakh Profile*, *supra* note 24.

³³ See Michael J. Bazylar & Rajika L. Shah, *The Unfinished Business of the Armenian Genocide: Armenian Property Restitution in American Courts*, 23 SW. J. INT'L L. 223, 227-28 and accompanying notes (2017).

³⁴ *Id.*

³⁵ See also CARLEY, *supra* note 27.

“unsupportive of the [Nagorno-Karabakh] peace process. . . .”³⁶ A 2019 report on social media manipulation by the Computational Propaganda Research Project at Oxford University concluded that Azerbaijan and Turkey were both “authoritarian countries deploying computational propaganda . . . as a tool of information control.”³⁷ Such control was expressed “in three distinct ways: to suppress fundamental human rights, discredit political opponents, and drown out dissenting opinions.”³⁸ The same report labeled Azerbaijan and Turkey as having “medium cyber troop capacity,” meaning they possessed full-time staff who coordinated with multiple actors, tools, and strategies for social media manipulation, including potentially abroad.³⁹

The July 2020 clashes gave rise to the first wave of heavily manipulated, pro-Azerbaijani social media, with a small group of accounts being responsible for a significant portion of the information shared.⁴⁰ Initial signs of the Azerbaijani government’s role in promoting this online activity quickly became apparent, as pro-regime student groups were some of the primary accounts engaging in online disinformation campaigns.⁴¹

By September 2020, when war broke out, the Azerbaijani government ramped up its attempts to control the online sphere, blocking or slowing access to most social media platforms while

³⁶ Azru Geybullayeva, *Azerbaijani Blogs Talk About Armenians: Introducing Hate 2.0*, OSSERVATORIO BALCANI E CAUCASO TRANSEUROPA (Feb. 3, 2012), <https://www.balcanicaucaso.org/eng/Areas/Azerbaijan/Azerbaijani-blogs-talk-about-Armenians-introducing-Hate-2.0-111320>.

³⁷ SAMANTHA BRADSHAW & PHILIP N. HOWARD, UNIV. OF OXFORD: OXFORD INT. INST., 2019 GLOBAL INVENTORY OF ORGANISED SOCIAL MEDIA MANIPULATION 5 (2019), <https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf>. The report defined computational propaganda as “the use of algorithms, automation, and big data to shape public life.” *Id.* at 1.

³⁸ *Id.*; see also Azru Geybulla, *In the Crosshairs of Azerbaijan’s Patriotic Trolls*, OPENDEMOCRACY (Nov. 22, 2016), <https://www.opendemocracy.net/en/odr/azerbaijan-patriotic-trolls/>.

³⁹ BRADSHAW & HOWARD, *supra* note 37. Armenia was listed as having “minimal cyber troop teams” that applied “a few tools of computational propaganda to a small number of platforms” and no foreign operations.

⁴⁰ Zarine Kharazian, *Patriotic Astroturfing in the Azerbaijan-Armenia Twitter War*, DFRLAB (July 21, 2020), <https://medium.com/dfrlab/patriotic-astroturfing-in-the-azerbaijan-armenia-twitter-war-9d234206cdd7>.

⁴¹ *Id.*

leaving Twitter mostly unblocked.⁴² This prompted a surge of virtual private network (VPN) app downloads in Azerbaijan as citizens tried to circumvent the block.⁴³ Although Twitter was not widely used in Azerbaijan, it ultimately helped the regime achieve its goals by allowing for greater surveillance and control of online information and providing additional channels for coordinating propaganda and harassment campaigns.⁴⁴

With the social media block in place, the pro-Azerbaijani content shared on social media platforms in the initial days of the Azerbaijani offensive in September originated mostly in countries friendly to Azerbaijan. A “substantial proportion” of such content shared in English was linked to accounts from Turkey and Pakistan.⁴⁵ Even online Turkish communities dedicated to sharing content about K-pop music mobilized to spread anti-Armenian hashtags.⁴⁶

The support for unadulterated violent rhetoric garnered on social media led to more displays of violent action being shared and broadcasted on various platforms. In Lyon, France, the Turkish ultranationalist militant group “Grey Wolves,” which is banned in a number of countries, posted videos of themselves on social media marching through neighborhoods with captions such as “looking for Armenians.”⁴⁷ The French police had to forcibly disperse the violent

⁴² Katy Pearce, *While Armenia and Azerbaijan Fought Over Nagorno-Karabakh, Their Citizens Battled on Social Media*, WASH. POST (Dec. 4, 2020), <https://www.washingtonpost.com/politics/2020/12/04/while-armenia-azerbaijan-fought-over-nagorno-karabakh-their-citizens-battled-social-media/>.

⁴³ An Azerbaijani Journalist, *Azerbaijanis Take Up Virtual Arms in Global Information War with Armenia*, EURASIANET (Oct. 11, 2020), <https://eurasianet.org/azerbaijanis-take-up-virtual-arms-in-global-information-war-with-armenia>.

⁴⁴ Pearce, *supra* note 42.

⁴⁵ ELISE THOMAS & ALBERT ZHANG, AUSTRALIAN STRATEGIC POL’Y INST., SNAPSHOT OF A SHADOW WAR: A PRELIMINARY ANALYSIS OF TWITTER ACTIVITY LINKED TO THE AZERBAIJAN-ARMENIA CONFLICT 20 (2020), <https://s3-ap-southeast-2.amazonaws.com/ad-aspi/2020-10/Snapshot%20of%20a%20shadow%20war.pdf>. Turkey, Pakistan, and Azerbaijan have launched coordinated hacking and social media campaigns in support of one another in previous conflicts.

⁴⁶ Lukas Andriukaitis, *Turkish Pop Culture Twitter Accounts Mobilize to Support Azerbaijan*, DRFLAB (Dec. 15, 2020), <https://medium.com/dfrlab/turkish-pop-culture-twitter-accounts-mobilize-to-support-azerbaijan-5b740511d792>.

⁴⁷ Hume, *supra* note 20.

mob who yelled threats that they were “going to kill the Armenians.”⁴⁸ Organized efforts of copy-pasted content targeted celebrities that showed any signs of support for Armenia, leading some, like rapper Cardi B, to retract their messages in the face of such overwhelming spam.⁴⁹

Perhaps most upsetting, videos and photographs apparently depicting war crimes and the brutal mistreatment of Armenian prisoners of war (POWs)—many of which seemed to have been filmed and posted by the alleged perpetrators themselves—were also widely circulated on social media.⁵⁰ A Human Rights Watch report noted that it was “telling that some of the [Azerbaijani] servicemen who carried out these abuses had no qualms about being filmed,” implying that the perpetrators feared no repercussions from the Azerbaijani regime for their crimes and that they felt emboldened to openly share their actions on social media platforms.⁵¹

B. THE RESPONSE FROM SOCIAL MEDIA COMPANIES WAS SLOW AND INEFFECTIVE

All of the social media companies that served as the main conduits for hate-based content—Instagram, TikTok, Facebook, Reddit, and Twitter—had policies in place at the time concerning hate speech and posts that incited, glorified, or otherwise served to spread violence. Despite that, activity violating such guidelines was rarely addressed effectively.

While some social media companies, such as Instagram, took some action to block, take down, or stop the spread of such information posted to their sites in connection with the 2020 Conflict, others, such as Facebook, TikTok, and Twitter, did not act or acted too late to allow for effective implementation of their anti-violence policies. For example, it took over a year of advocacy and the leaking during the war of an internal memo that exposed Facebook’s failures before

⁴⁸ *Id.*

⁴⁹ *See, e.g.,* @josh_emerson, TWITTER (now X) (Oct. 6, 2020, 3:54 AM), https://twitter.com/josh_emerson/status/1313432532487208962 (posting a screengrab of the social media campaign targeting Cardi B).

⁵⁰ *See, e.g.,* *Azerbaijan: Armenian Prisoners of War Badly Mistreated*, HUM. RTS. WATCH (Dec. 2, 2020), <https://www.hrw.org/news/2020/12/02/azerbaijan-armenian-prisoners-war-badly-mistreated>.

⁵¹ *Id.*

Facebook finally took down thousands of accounts and pages tied to the Azerbaijani regime that targeted opposition figures and independent media.⁵² Less than six months later, those troll networks returned to the platform and launched further harassment campaigns.⁵³

IV. INTERNATIONAL CRIMINAL TRIBUNALS HAVE HELD COMPANIES AND MEDIA EXECUTIVES LIABLE FOR THEIR ROLE IN ATROCITY CRIMES

⁵² Craig Silverman & Ryan Mac, *It Took Facebook More Than a Year—And a Whistleblower—To Remove Troll Farm Connected To Azerbaijan’s Ruling Party*, BUZZFEED NEWS (Oct. 8, 2020, 9:43 AM), <https://www.buzzfeednews.com/article/craigsilverman/facebook-azerbaijan-troll-farm>; Craig Silverman et al., *‘I Have Blood on My Hands’: A Whistleblower Says Facebook Ignored Global Political Manipulation*, BUZZFEED NEWS (Sept. 14, 2020, 12:36 PM), <https://www.buzzfeednews.com/article/craigsilverman/facebook-ignore-political-manipulation-whistleblower-memo>.

⁵³ Julia Carrie Wong & Luke Harding, *‘Facebook Isn’t Interested In Countries Like Ours’: Azerbaijan Troll Network Returns Months After Ban*, GUARDIAN (Apr. 13, 2021), <https://www.theguardian.com/technology/2021/apr/13/facebook-azerbaijan-ilham-aliyev>. These manipulative online practices are known as “astroturfing.” Astroturfing often entails using online identities, including fake groups and accounts, to “create an impression of widespread grassroots support for a policy, individual, or product, where little such supports exists.” Adam Bienkob, *Astroturfing: What is it and Why does it Matter?*, GUARDIAN (Feb. 8, 2012), <https://www.theguardian.com/commentisfree/2012/feb/08/what-is-astroturfing>. “Ephemeral astroturfing” is a quick and coordinated campaign in which an account is created for the purpose of creating a social media trend, while simultaneously quickly deleting content containing keywords relating to those same trends. See Tuğrulcan Elmas et al., *Ephemeral Astroturfing Attacks: The Case of Fake Twitter Trends*, in 2021 IEEE EUROPEAN SYMPOSIUM ON SECURITY AND PRIVACY (EUROS&P), 403, 403-05 (2021), <https://doi.ieeecomputersociety.org/10.1109/EuroSP51992.2021.00035>. Such measures leave little evidence behind once their goal of “trending” a hashtag or topic is successful. *Id.*

Azerbaijan and Turkey have been observed using such tactics, especially on Twitter. *Id.* Perhaps not coincidentally, Twitter expanded its hate speech policies in the fall of 2020, leading to over 1.1 million different accounts facing action and over 3.8 million tweets being removed. Kurt Wagner/Bloomberg, *Twitter Penalizes Record Number of Accounts for Posting Hate Speech*, TIME (July 14, 2021), <https://time.com/6080324/twitter-hate-speech-penalties/>. (Nonetheless, thousands of accounts dedicated to launching harassment campaigns, spreading propaganda and genocide denial, and promoting violence remained active.)

Discrimination-based atrocity crimes, including persecution as a crime against humanity and genocide, require the spread of hate speech and disinformation to lay the ideological groundwork of violence and destruction.⁵⁴ While disinformation and misinformation remain more nebulous concepts, they also work to normalize the dehumanization of a group, in order to validate the group's victimization.⁵⁵ Consequently, mass media plays an integral role in facilitating atrocity crimes by enabling the weaponization of language to engender fear and mobilize a destructive response.⁵⁶

Since the end of World War II, international criminal tribunals ("ICTs") have recognized this entanglement between the media and atrocities. Multiple ICTs have imputed liability to media company executives for the spread of hateful and inflammatory messages on their platforms that catalyzed the commission of atrocity crimes—particularly the crime that we now know as direct and public incitement to genocide. This Section reviews those precedents to identify the circumstances in which liability may be found.

A. THE NUREMBERG IMT FOUND LIABILITY WHERE A NEWSPAPER PUBLISHER CONTINUED TO PUBLISH ARTICLES INCITING GENOCIDAL VIOLENCE WHILE AWARE OF THE STATE'S GENOCIDAL VIOLENCE

Precedent from the International Military Tribunal at Nuremberg ("IMT"), the first ICT, lays the foundation for modern international criminal law. At the IMT, because neither the crime of incitement to genocide nor even genocide had yet been defined, incitement to "murder and extermination" was charged as a form of persecution under the umbrella of crimes against humanity. IMT jurisprudence also laid the path for what became the modern crime of direct and public incitement (of others) to commit genocide.⁵⁷

On October 18, 1945, the prosecutors at the IMT indicted 24 leading Nazi officials. Each defendant was charged with at least one

⁵⁴ Frank Chalk, *Intervening to Prevent Genocidal Violence: The Role of the Media*, in *THE MEDIA AND THE RWANDA GENOCIDE* 375, 375-80 (Allan Thompson ed., 2007).

⁵⁵ *Id.*

⁵⁶ *Id.* at 376.

⁵⁷ *See, e.g.*, Rome Statute of the International Criminal Court, art. 25(e), July 17, 1998, 2187 U.N.T.S. 90.

of the four possible crimes: (1) crimes against peace (the modern crime of aggression), (2) war crimes, (3) crimes against humanity, or (4) common plan or conspiracy to commit [crimes against peace, war crimes, or crimes against humanity].⁵⁸

One of the 24 individuals charged was Julius Streicher. Streicher was the publisher of *Der Stürmer*, “an anti-Semitic German weekly newspaper” published from 1923 to 1945; he was also the editor until 1933.⁵⁹ Widely known as the “Jew-Baiter Number One,” in his capacity at the publication, Streicher heralded a “call for the annihilation of the Jewish race.”⁶⁰ Twenty-three articles in *Der Stürmer* explicitly called for the “root and branch” extermination of Jewish people,⁶¹ urging that “only when world Jewry had been annihilated would the Jewish problem be solved.”⁶² Dehumanizing phrases used in reference to Jewish people, such as “germ,” “pest,” and “parasite . . . who must be destroyed in the interest of mankind,” were commonplace in *Der Stürmer* articles.⁶³

Though Streicher claimed he strived solely to ostracize Jews as “aliens” and facilitate their deportation rather than death and denied having knowledge of the mass extermination of Jews, the tribunal placed little weight on Streicher’s testimony.⁶⁴ Rather, according to the court, Streicher actually intensified his campaign against the Jewish people when he gained “knowledge of the extermination of the Jews in the Occupied Eastern Territory.”⁶⁵ The court noted evidence that made “it clear that he continually received current information on the progress of the ‘final solution.’”⁶⁶ For example, *Der Stürmer*’s press photographer was sent to visit the ghettos in the spring of 1943. Also, Streicher received and read another newspaper which “carried in each issue accounts of Jewish atrocities.”⁶⁷

⁵⁸ *The Trial of German Major War Criminals: Proceedings of the International Military Tribunal Sitting at Nuremberg, Germany*, Vol. 1, at 28 (1946).

⁵⁹ *Id.* at 301.

⁶⁰ *Id.*

⁶¹ *Id.* at 302.

⁶² *Id.*

⁶³ *Id.* at 301.

⁶⁴ *Id.* at 304.

⁶⁵ *The Trial of German Major War Criminals: Proceedings of The International Military Tribunal Sitting at Nuremberg, Germany*, Part 22, at 301 (1946).

⁶⁶ *Id.*

⁶⁷ *Id.*

In light of this evidence, the court determined that Streicher “infected the German mind with the virus of anti-Semitism and incited the German people to active persecution” via the widespread publication of *Der Stürmer*.⁶⁸ Accordingly, the Court held that “Streicher’s incitement to murder and extermination at the time when Jews in the East were being killed . . . constitute[d] a crime against humanity.”⁶⁹

In contrast, the IMT acquitted another defendant, Hans Fritzsche, the Head of the Radio Section of the German Propaganda Ministry. Though Fritzsche shared Streicher’s rampant anti-Semitism, his broadcasts were found not to have “urge[d] persecution or extermination of Jews,” and there was “no evidence that he was aware of their extermination in the East.”⁷⁰ Moreover, Fritzsche appeared to have attempted to temper Streicher’s hateful diatribe as he “twice attempted to have publication of the anti-Semitic *Der Stürmer* suppressed, though unsuccessfully.”⁷¹ In acquitting Fritzsche, the court emphasized the significance of the lack of language impelling extermination and knowledge of the atrocities being committed.⁷²

The difference in outcomes between *Streicher* and *Fritzsche* establishes the idea that media executives may avoid liability if they are not directly advocating for the extermination of a particular group of people or genuinely (and reasonably) lack awareness of atrocities being committed against that group. It also serves as a warning that, in the context of a particularly volatile conflict environment, media executives must be careful to avoid adding fuel to the fire.

B. THE SUBSEQUENT NUREMBERG TRIALS FOUND SECONDARY CORPORATE LIABILITY WHERE A COMPANY KNOWINGLY SUPPLIED A COMMODITY TO THE STATE WHILE AWARE THE STATE WAS USING THE COMMODITY TO COMMIT SERIOUS CRIMES

⁶⁸ *Id.*

⁶⁹ *Id.*

⁷⁰ *Id.*

⁷¹ *Id.*

⁷² *Id.* at 338.

In addition to the IMT, a number of other trials were held focusing on determining the degree to which civil and military society aided and abetted the Nazis' Final Solution. One of these cases, *United Kingdom v. Tesch*, highlights the liability of owners of corporations that provide the means to the end of genocide or mass atrocities.

In *Tesch*, the main question centered on the liability of senior executives at a company that distributed Zyklon B,⁷³ the gas used to murder prisoners in extermination camps throughout the Third Reich. The first defendant in the case, Bruno Tesch, sold Zyklon B through his firm, Tesch and Stabenow.⁷⁴ Karl Weinbacher, the second defendant, was Tesch's second-in-command.⁷⁵ The third defendant, Joachim Drosihn, was the firm's gassing technician.⁷⁶ The British Military Court charged all three men with the war crime of "supply[ing] poison gas used for extermination of [individuals] interned in concentrations camps well knowing that the said gas was to be so used."⁷⁷

A core question in any such case is whether the defendant acted with the requisite mens rea or mental element.⁷⁸ In *Tesch*, according to

⁷³ Zyklon B is a highly poisonous insecticide originally intended for use against rats. When exposed to air, Zyklon B pellets convert into a lethal gas. Leaders of Nazi Germany determined this was the most efficient way to kill prisoners, which led to mass murder at many extermination camps. *At the Killing Centers*, U.S. HOLOCAUST MEM'L MUSEUM, <https://encyclopedia.ushmm.org/content/en/article/at-the-killing-centers> (last edited Mar. 3, 2023).

⁷⁴ *United Kingdom v. Tesch (The Zyklon B Case)*, Case No. 9, 1 Law Rep. Trials War Crim. 93 (Brit. Mil. Ct. Hamburg, Mar. 1-8, 1946).

⁷⁵ *Id.*

⁷⁶ *Id.*

⁷⁷ *Id.*

⁷⁸ Modernly, at the International Criminal Court, the Rome Statute requires that, unless otherwise provided, the material elements of a crime must be committed with intent and knowledge. Rome Statute of the International Criminal Court, art. 30(1), *supra* note 57, 2187 U.N.T.S. at 90. Intent is satisfied when "(a) In relation to conduct, that person means to engage in the conduct; (b) In relation to a consequence, that person means to cause that consequence or is aware that it will occur in the ordinary course of events." *Id.* art. 30(2). Knowledge "means awareness that a circumstance exists or a consequence will occur in the ordinary course of events." *Id.* art. 30(3). At least one author has argued that the holding in *Tesch* can be applied to the senior executives of a social media company if the company provided a platform to individuals carrying out a propaganda campaign of incitement and the executive knew or should have known the platform directly

the prosecution, knowingly supplying “a commodity to . . . the State which was using that commodity for the mass extermination of [civilians] was a war crime,” and thus, liability attached to any individual who “put the means to commit the crime into the hands of those who actually carried it out.”⁷⁹

Prosecution witnesses, including bookkeepers and stenographers at the company, testified that Tesch and Weinbacher were aware of the lethal use of Zyklon B in the extermination camps and yet “continued to arrange supplies of gas to” the camps “in ever-increasing quantities.”⁸⁰ Therefore, Tesch and Weinbacher knew the SS was using Zyklon B to exterminate civilians in extermination camps. Tesch and Weinbacher, however, argued that since they were not present at the concentration camps nor did they personally place the Zyklon B pellets in the gas chambers, they were not liable for the crimes charged.⁸¹ They also highlighted Zyklon B’s non-lethal purpose of delousing the camps’ quarters, arguing that any increase in the SS purchase order was due to the increase of prisoners in the camps.⁸²

The Court concluded that both Tesch and Weinbacher were “competent business men.”⁸³ Given the German public knowledge by at least 1943 that Zyklon B was “being used for killing people,”⁸⁴ the Court concluded that Tesch and Weinbacher knew or should have known the SS was using extra shipments of Zyklon B to extermination camps not for delousing but rather as a weapon of mass murder.⁸⁵ Accordingly, the Court found Tesch and Weinbacher guilty and sentenced both to execution.⁸⁶

Drosihn, on the other hand, was acquitted due to his lack of influence over the transfer of gas to the camps and, therefore, his

assisted the incitement. Neema Hakim, Comment, *How Social Media Companies Could Be Complicit in Incitement to Genocide*, 21 CHI. J. INT’L L. 83, 111 (2020).

⁷⁹ *Tesch* (The Zyklon B Case), Case No. 9, 1 Law Rep. Trials War Crim. at 94.

⁸⁰ *Id.* at 94-95.

⁸¹ *Id.* at 97.

⁸² *Id.*

⁸³ *Id.* at 101.

⁸⁴ *Id.* at 96.

⁸⁵ *Id.* at 101.

⁸⁶ *Id.* at 102.

inability to prevent it.⁸⁷ Drosihn had a “subordinate position” in the company in relation to his limited knowledge and influence over the “firm’s business activities.”⁸⁸ Further, Drosihn spent a majority of the year traveling; when Tesch and Weinbacher were traveling and Drosihn was at company headquarters, he did not have “the power of attorney.”⁸⁹ Ultimately, the Court concluded Drosihn was not in a position at the firm “to influence the transfer of gas to Auschwitz or prevent it.”⁹⁰

C. THE ICTR’S MEDIA CASE HELD MEDIA COMPANIES EXECUTIVES LIABLE FOR INCITING GENOCIDE DUE TO THE MESSAGING DISSEMINATED ON THEIR PLATFORMS

Building upon the precedent set at the IMT and NMT, the International Criminal Tribunal for Rwanda (“ICTR”) was the first modern ICT to examine the role of mass media in facilitating atrocity crimes in what is known as the “Media Case.”⁹¹

1. *RTLM Radio Broadcasts Advocated for the Extermination of Tutsis*

Ferdinand Nahimana and Jean-Bosco Barayagwiza were founders of the Rwandan media organization Radio-Television Libre des Mille Collines (RTLM).⁹² Nahimana was viewed as the founder and director, while Barayagwiza was the second in command.⁹³ Both Nahimana and Barayagwiza remained in the top management of RTLM and represented the radio at the highest level in meetings with

⁸⁷ *Id.*

⁸⁸ *Id.* at 100.

⁸⁹ *Id.*

⁹⁰ *Id.* at 102.

⁹¹ Recent Case, Prosecutor v. Nahimana, Barayagwiza, and Ngeze (Media Case), Case No. ICTR-99-52-T (*Int’l Crim. Trib. for Rwanda Trial Chamber I Dec. 3, 2003*), 117 HARV. L. REV. 2769, 2769 (2004).

⁹² Prosecutor v. Nahimana (Media Case), Case No. ICTR-99-52-T, Trial Chamber Judgement, ¶ 567 (Dec. 3, 2003).

⁹³ *Id.*

the Rwandan Ministry of Information.⁹⁴ They also controlled the finances and were members of the board of directors of RTLM.⁹⁵

In the 1980s, the Rwandan government had subsidized the production of radios, which were then sold at a reduced price or given to those in the administrative structure of the party.⁹⁶ As a result, the radio became an increasingly important source of information to the Rwandan public, as well as an avenue for entertainment and a focus of social life.⁹⁷ RTLM started broadcasting in July 1993 and quickly gained popularity. Many people, particularly younger generations, were seen listening to RTLM on the streets and at work or playing it in bars, taxis, and markets.⁹⁸

Prior to April 1994, RTLM's broadcasts primarily discussed ethnicity in the context of the nation's history and the politics of Hutu-Tutsi relations, promoted as attempts to "raise awareness" about these issues.⁹⁹ However, these broadcasts soon devolved into "ethnic stereotyping in economic terms as well as political,"¹⁰⁰ such as distorting facts to portray the Tutsi as unjustifiably wealthy in a country of enormous poverty—a tactic also utilized by Nazi Germany to target Jews—as well as "ethnic stereotyping in reference to physical characteristics,"¹⁰¹ thus "contributing to increasing hostility against the Tutsi."¹⁰²

A month before the onset of the genocide in April 1994, the station began to arbitrarily identify various Tutsi individuals as "security risks" and warned listeners to "rise up."¹⁰³ The Trial Chamber found these depictions "heated up heads;"¹⁰⁴ they "promoted contempt and hatred for the Tutsi population and called listeners to seek out and take up arms against the enemy."¹⁰⁵ As one witness observed, "[w]hat RTLM did was almost to pour petrol – to spread

⁹⁴ *Id.* at ¶ 970.

⁹⁵ *Id.*

⁹⁶ *Id.* at ¶ 342.

⁹⁷ *Id.*

⁹⁸ *Id.*

⁹⁹ *Id.* at ¶ 345.

¹⁰⁰ *Id.* at ¶ 363.

¹⁰¹ *Id.* at ¶ 368.

¹⁰² *Id.* at ¶ 365.

¹⁰³ *Id.* at ¶ 371, 375.

¹⁰⁴ *Id.* at ¶ 371.

¹⁰⁵ *Id.* at ¶ 486.

petrol throughout the country little by little so that one day it would be able to set fire to the whole country.”¹⁰⁶

After April 6, 1994, with the country aflame with ethnic hostilities, the “virulence and the intensity of RTLM broadcasts propagating ethnic hatred and calling for violence increased.”¹⁰⁷ RTLM’s programs escalated its anti-Tutsi rhetoric by unequivocally “defin[ing] the enemy as the Tutsi”¹⁰⁸ and “explicitly call[ing] for the extermination of the Tutsi ethnic group.”¹⁰⁹ One transmission even went so far as to “describe the physical characteristics of the ethnic group as a guide to selecting targets of violence,”¹¹⁰ explaining “the reason we will exterminate them is that they belong to one ethnic group”—the Tutsis.¹¹¹ Another lauded the idea of “exterminating the Tutsi from the surface of the earth . . . to make them disappear for good.”¹¹²

As a result of its prevalence throughout the region, the RTLM radio station was one of the main sources of mass media propaganda.¹¹³ More than merely stoking the flames of discord, RTLM was described as “constantly asking people to kill other people, to look for those who were in hiding, and to describe the hiding places of those who were described as being accomplices”¹¹⁴ and even effectuated targeted killings against specific individuals by “publishing lists of Tutsi names and asking for people to come forth and provide information on those listed.”¹¹⁵ RTLM broadcasts thus “relentlessly sen[t] the message that the Tutsi were the enemy and had to be eliminated once and for all.”¹¹⁶

As widescale bloodshed erupted, written complaints and notices of violations were sent to Ferdinand Nahimana and Jean-Bosco

¹⁰⁶ *Id.* at ¶ 436.

¹⁰⁷ *Id.* at ¶ 486 (April 6, 1994, marks the date Rwandan President Juvénal Habyarimana was assassinated, which served as the catalyst for the Rwandan Genocide).

¹⁰⁸ *Id.* at ¶ 392.

¹⁰⁹ *Id.* at ¶ 486.

¹¹⁰ *Id.* at ¶ 396.

¹¹¹ *Id.*

¹¹² *Id.* at ¶ 483.

¹¹³ *Id.* at ¶ 488 (“Radio was the medium of mass communication with the broadest reach in Rwanda”).

¹¹⁴ *Id.*

¹¹⁵ *Id.* at ¶ 487.

¹¹⁶ *Id.* at ¶ 488.

Barayawiza; however, despite participating in meetings with the Ministry of Information and receiving injunctions, RTLM ignored the government's call to end broadcasts of this type and continued to promote violence.¹¹⁷

2. *Kangura Newspaper Articles Similarly Demonized Tutsis*

Hassan Ngeze was a journalist by trade and in 1990 founded the newspaper *Kangura*, where he was Editor-in-Chief for the entirety of its existence.¹¹⁸ As such, Ngeze was responsible for the “overall direction of the paper” and “all authority connected with the newspaper remained in his hands.”¹¹⁹ Depending on sales, about 1,500 to 3,000 copies of each issue were printed.¹²⁰ *Kangura* was the most well-known newspaper in Rwanda at the time.¹²¹

Ngeze often wrote articles for *Kangura* himself.¹²² Beginning in 1991, as per a requirement of the Kigali prosecutor, a notice was printed on the bottom of the cover page of every issue stating that “the content of the articles binds the author and the publisher.”¹²³ Though the editorial team met to discuss each issue, Ngeze was the ultimate authority and had the last word as to what was published.¹²⁴ Thus, Ngeze “controlled the publication and was responsible for its contents.”¹²⁵

Many articles in *Kangura* portrayed Tutsis as a group as “the enemy, as evil, dishonest and ambitious.”¹²⁶ One of the most infamous was the “Ten Commandments,” published in December 1990.¹²⁷ Addressed to what it called the Hutu majority, the article exhorted Hutus to “wake up!” and “[t]ake all necessary measures to deter [Tutsi] from launching a fresh attack” because “the enemy” was waiting for a

¹¹⁷ *Id.*

¹¹⁸ *Id.* at ¶ 122, 123.

¹¹⁹ *Id.* at ¶ 123.

¹²⁰ *Id.* at ¶ 122.

¹²¹ *Id.*

¹²² *Id.* at ¶ 129.

¹²³ *Id.* at ¶ 123.

¹²⁴ *Id.* at ¶ 129.

¹²⁵ *Id.* at ¶ 135.

¹²⁶ *Id.* at ¶ 152-59.

¹²⁷ *Id.* at ¶ 138.

“more propitious moment[] to decimate us.”¹²⁸ It described the Tutsi as “bloodthirsty” and raised the specter of “Tutsi domination over the Hutu.”¹²⁹ The article further claimed that Tutsi women were intentionally married or sold to Hutu intellectuals and high-placed Hutu officials in order to “serve as spies.”¹³⁰ The article then urged Hutus to “become aware of a new Hutu ideology,” “cease feeling pity for the Tutsi,” and follow the ten commandments.¹³¹ Those who did not were explicitly labeled as traitors.¹³²

3. *The ICTR Trial Chamber Found Individual Criminal Responsibility for Media Company Heads Who Fanned the Flames of Violence*

In 2003, ICTR Trial Chamber I found Ferdinand Nahimana, Jean-Bosco Barayagwiza, and Hassan Ngeze each guilty of genocide; conspiracy to commit genocide; direct and public incitement to commit genocide; and persecution and extermination as crimes against humanity.¹³³

The Trial Chamber held that, as the “number one” and “number two” of RTLM’s top management, Nahimana and Barayagwiza had a duty to “take necessary and reasonable measures to prevent the killing of Tutsi civilians instigated by the RTLM.”¹³⁴ Given that both defendants had been on notice about the alarming amplification of antagonism in the RTLM’s messaging, the Trial Chamber found they “knew what was happening at RTLM” and “failed to exercise the authority vested in them . . . to prevent the genocidal harm that was caused by RTLM programming.”¹³⁵ This omission factored into the Chamber’s decision to hold them criminally responsible.¹³⁶

With respect to *Kangura*, the Chamber found that, by publishing articles and editorials that “conveyed contempt and hatred for the Tutsi ethnic group, and for Tutsi women in particular as enemy

¹²⁸ *Id.*

¹²⁹ *Id.* at ¶ 139.

¹³⁰ *Id.* at ¶ 139.

¹³¹ *Id.* at ¶ 139.

¹³² *Id.*

¹³³ *Id.* at ¶¶ 1092, 1093, 1094.

¹³⁴ *Id.* at ¶ 973.

¹³⁵ *Id.* at ¶ 970.

¹³⁶ *Id.* at ¶¶ 973-74.

agents,” and specifically in publishing the “Ten Commandments,” *Kangura* “fan[ned] the flames of ethnic hatred, resentment and fear against the Tutsi population.”¹³⁷ Moreover, *Kangura* promoted violence against Tutsis through fear-mongering and hate propaganda, calling on readers to “take all necessary measures to stop the enemy.”¹³⁸ Thus, *Kangura* “paved the way for genocide in Rwanda” against the Tutsi population by “whipping the Hutu population into a killing frenzy.”¹³⁹

Importantly, the Trial Chamber recognized that “the power of the media to create and destroy fundamental human values comes with great responsibility,” and thus, “those who control such media are accountable for its consequences.”¹⁴⁰ The Trial Chamber concluded that the defendants bore individual criminal responsibility stemming from their “ownership and institutional control over the media”¹⁴¹ and their use of it “for the collective communications of ideas and for the mobilization of the population on a grand scale.”¹⁴²

Furthermore, the Chamber noted that even when parroting the messaging of others through their outlets, editors, and publishers have “generally been held responsible for media they control.”¹⁴³ Publishers and editors are “regarded as equally responsible” for the words of others that they distribute on their platforms on the grounds that they are providing a forum and that as owners they have “the power to share the editorial direction.”¹⁴⁴ A publisher’s or editor’s intent, specifically whether or not the purpose of publicly transmitting the material was in good faith or part of a campaign of malice, determined the scope of this responsibility.¹⁴⁵

Conversely, the key factor the Chamber identified in absolving publishers and editors of this liability was whether they maintained a critical distance from the published content, such as offering

¹³⁷ *Id.* at ¶ 950.

¹³⁸ *Id.*

¹³⁹ *Id.*

¹⁴⁰ *Id.* at ¶ 945.

¹⁴¹ *Id.* at ¶ 979.

¹⁴² *Id.*

¹⁴³ *Id.* at ¶ 1001.

¹⁴⁴ *Id.* at ¶ 1003.

¹⁴⁵ *Id.*

disclaimers or opposing points of view.¹⁴⁶ The Chamber held this “clear distancing” was crucial in cases where the “disseminated views constitute ethnic hatred and call to violence” to “avoid conveying an endorsement of the message.”¹⁴⁷ As such, the Chamber rejected the defendants’ claims that some of the statements published in the broadcast of RTLM or *Kangura* were simply facts and informational in their nature.¹⁴⁸

4. *The ICTR Appeals Chamber Affirmed Media Executives’ Responsibility to Prevent the Spread of Violent Content*

Each defendant appealed, and in 2007, the ICTR Appeals Chamber issued its final judgment in the Media Case.

Regarding the charge of genocide, the Chamber found that in some cases there was insufficient evidence to conclude that RTLM broadcasts listing names of certain Tutsis substantially contributed to their murder, either because the murders themselves were not sufficiently established or because there were intervening causes.¹⁴⁹ The Chamber found that it was not established beyond a reasonable doubt that Nahimana “played an active role in the [RTLM] broadcasts instigating the commission of genocide” after the genocide began on April 6, 1994, and that there was insufficient proof the editorials and other texts Nahimana allegedly asked to be read out on air instigated the killing of Tutsis.¹⁵⁰ The Chamber also found that there was insufficient evidence to demonstrate that Barayagwiza “continued to exercise effective control over RTLM after” April 6, 1994, particularly because he was only second in command.¹⁵¹ Finally, it had not been proven beyond a reasonable doubt that *Kangura* “substantially contributed to” the genocide, even though there was “probably a link”

¹⁴⁶ *Id.*; see also *id.* at ¶ 992 (citing *Jersild v. Denmark*, App. no. 15890/89 (Eur. Ct. Human Rts. Sept. 23, 1994), where the European Court of Human Rights overturned a journalist’s conviction under a Danish law prohibiting discrimination despite interviewing a racist youth group who propagated hate speech because the journalist “clearly disassociated himself from the persons interviewed”).

¹⁴⁷ *Id.*

¹⁴⁸ *Id.* at ¶ 1024.

¹⁴⁹ *Prosecutor v. Nahimana*, Case No. ICTR-99-52-A, Appeals Chamber Judgement, ¶¶ 507-13 (Nov. 28, 2007).

¹⁵⁰ *Id.* at ¶¶ 596-98.

¹⁵¹ *Id.* at ¶ 635.

between Ngeze's acts and the genocide "owing to the climate of violence to which the publication contributed and the incendiary discourse it contained," and thus Ngeze could not be found guilty of genocide.¹⁵²

Regarding the charge of direct and public incitement to genocide, the Appeals Chamber engaged in a lengthy discussion. The Chamber first distinguished *instigation* of genocide (a mode of responsibility applicable to any of the crimes chargeable under the tribunal's Statute, in which the accused incurred individual criminal responsibility if the instigation "in fact substantially contributed to the commission of" genocide, crimes against humanity, or war crimes) with *direct and public incitement to genocide* (a crime in and of itself, which was punishable as an inchoate offense even if no act of genocide occurred).¹⁵³

Additionally, while the meaning of "public" was fairly clear, the meaning of "direct" required greater explanation.¹⁵⁴ In the Chamber's view, there was a difference between "hate speech in general (or speech inciting discrimination or violence) and direct and public incitement to commit genocide."¹⁵⁵ Direct incitement "assume[d] that the speech [was] a direct appeal to commit" one of the *actus reus* of genocide, something "more than a mere vague or indirect suggestion."¹⁵⁶ Thus, hate speech that "[did] not directly call for the commission of genocide" would not rise to the level of direct and public incitement.¹⁵⁷ Moreover, the specific "acts constituting direct and public incitement to commit genocide must be clearly identified."¹⁵⁸

However, the Appeals Chamber did confirm that "the Trial Chamber did not alter the constituent elements of the crime of direct and public incitement to commit genocide in the media context," referencing the Trial Chamber's review of international precedent, including *Streicher* and *Fritzsche* (as well as various human rights tribunal opinions).¹⁵⁹ Thus, the Appeals Chamber approved of the

¹⁵² *Id.* at ¶ 519.

¹⁵³ *Id.* at ¶ 678.

¹⁵⁴ *Id.* at ¶ 691.

¹⁵⁵ *Id.* at ¶ 692.

¹⁵⁶ *Id.*

¹⁵⁷ *Id.* at ¶¶ 692-93.

¹⁵⁸ *Id.* at ¶ 726.

¹⁵⁹ *Id.* at ¶¶ 693-95.

“broad guidelines for interpreting and characterizing media discourse” that the Trial Chamber articulated.¹⁶⁰

The Appeals Chamber also affirmed that “contextual elements” such as local culture and linguistic nuance, and the author’s political and community affiliation, were relevant in determining whether speech constituted direct and public incitement to genocide.¹⁶¹ Where speech was potentially ambiguous in meaning, its “true message” was determined by “how a speech was understood by its intended audience.”¹⁶² If the message remained ambiguous even in context, it could not constitute a direct and public incitement to genocide.¹⁶³ It was not necessary that the speech “explicitly call[] for extermination” or be “entirely unambiguous for all types of audiences.”¹⁶⁴

The Appeals Chamber noted that “the purpose of the speech is indisputably a factor in determining whether there is direct and public incitement to commit genocide.”¹⁶⁵ Thus, “the mere fact that genocide occurred” following the speech in question was not necessarily sufficient to demonstrate that “individuals in control of the media intended to incite the commission of genocide,” because the genocide “could have been the result of other factors.”¹⁶⁶ As a result, it could not be “the only evidence adduced to conclude that the purpose of the speech (and of its author) was to incite” genocide.¹⁶⁷

In light of these principles, the Appeals Chamber affirmed the Trial Chamber’s holding that RTLM broadcasts after April 6, 1994 “called for the extermination of Tutsi and amounted to direct and public incitement to commit genocide.”¹⁶⁸ Additionally, several articles published in *Kangura* after April 6, 1994 contained direct calls for Hutu to “stand united in order to exterminate the Tutsi,” and thus constituted direct and public incitement to commit genocide.¹⁶⁹

¹⁶⁰ *Id.* at ¶ 695.

¹⁶¹ *Id.* at ¶¶ 697-98.

¹⁶² *Id.* at ¶¶ 700-01.

¹⁶³ *Id.* at ¶ 701.

¹⁶⁴ *Id.* at ¶ 702.

¹⁶⁵ *Id.* at ¶ 706.

¹⁶⁶ *Id.* at ¶ 709.

¹⁶⁷ *Id.*

¹⁶⁸ *Id.* at ¶¶ 757-58.

¹⁶⁹ *Id.* at ¶¶ 771-73.

Accordingly, the ICTR Appeals Chamber (1) affirmed Nahimana's convictions for direct and public incitement to genocide and persecution as a crime against humanity solely on the basis of superior responsibility and reversed all other convictions; (2) affirmed Barayagwiza's convictions for genocide, extermination as a crime against humanity, and persecution as a crime against humanity, all on the basis of individual criminal responsibility, and reversed all other charges; and (3) affirmed Ngeze's convictions for aiding and abetting genocide, direct and public incitement to genocide, and aiding and abetting extermination as a crime against humanity, all on the basis of individual criminal responsibility, and reversed all other convictions.¹⁷⁰

V. THE INTERNATIONAL COMMUNITY RECOGNIZED FACEBOOK'S ROLE IN EXACERBATING ETHNIC TENSIONS IN MYANMAR

While *Streicher*, *Fritzsche*, *Zyklon B*, and the ICTR Media Case set forth the leading international criminal precedents relating to media executives' primary and/or secondary liability for speech published on their platforms, more recent developments highlight the particular challenges for curtailing harmful speech that exist in the era of social media. Indeed, social media companies received a wake-up call when the Independent International Fact-Finding Mission on Myanmar (the "IIFFM") found in 2018 that Facebook played a key role in the incitement of violence against Rohingya Muslims.¹⁷¹

Rakhine state in northern Myanmar was historically comprised of two main groups: the Rakhine Buddhists and the Rohingya Muslims.¹⁷² Decades-long ethnic and religious tensions in Rakhine state had "often [been] ascribed to poor relations between the Rohingya and the Rakhine, reflective of deeply rooted grievances and prejudices."¹⁷³ Adding fuel to the fire, Myanmar military soldiers systematically oppressed and persecuted the Rohingya.¹⁷⁴

¹⁷⁰ *Id.* at ¶¶ 345-46.

¹⁷¹ Hum. Rts. Council, *Report of the Independent International Fact-Finding Mission on Myanmar*, U.N. Doc. A/HRC/39/64, at 4 (Sept. 2018) (emphasis added).

¹⁷² Hum. Rts. Council, *Report of the Detailed Findings of the Independent International Fact-Finding Mission on Myanmar*, U.N. Doc. A/HRC/39/CRP.2, at 339, 22 (Sept. 2018).

¹⁷³ *Id.* at 174.

¹⁷⁴ *Id.*

Following a series of small-scale attacks carried out in August 2017 in Rakhine state by the Arakan Rohingya Salvation Army (“ARSA”),¹⁷⁵ the Myanmar military (referred to as the Tatmadaw), in coordination with local Rakhine police, initiated a brutal and widespread campaign targeting Rohingya civilians. The “hallmarks of Tatmadaw operations”¹⁷⁶ included sexual violence and “exclusionary and discriminatory rhetoric.”¹⁷⁷ The Rohingya were subject to “[l]arge-scale massacres” where “[m]en, women and children were killed.”¹⁷⁸ The violence included burning people alive in huts, resulting in “[e]ntire villages [being] wiped off the map.”¹⁷⁹ Reports of “[m]ultiple victims with single gunshot wounds to the head” indicated that victims were subjected to “execution style killings” and later buried in “mass grave[s].”¹⁸⁰ Torture techniques of the Tatmadaw included “performing sexual violence, including rape,” as well as “making victims dig their ‘own’ graves.”¹⁸¹ As a result, in a matter of weeks nearly one million Rohingya fled across the border to Bangladesh seeking safety, resulting in “large-scale . . . displacement.”¹⁸² “The nature, scale and organization of the operation suggest[ed] a level of preplanning and design by the Tatmadaw leadership that was consistent with the vision of the Commander-in-Chief, Senior General Min Aung Hlaing.”¹⁸³

A. FACEBOOK PLAYED A CRUCIAL ROLE IN FACILITATING VIOLENCE AGAINST THE ROHINGYA

As a result of the violence in Myanmar, the UN Human Rights Council authorized the IFFM to establish “the facts and circumstances of the alleged recent human rights violations by military and security forces, and abuses, in Myanmar, in particular in Rakhine State.”¹⁸⁴ The

¹⁷⁵ *Id.* at 180.

¹⁷⁶ *Id.* at 346.

¹⁷⁷ *Id.* at 350.

¹⁷⁸ *Id.* at 347.

¹⁷⁹ *Id.* at 347.

¹⁸⁰ *Id.* at 40.

¹⁸¹ *Id.* at 48.

¹⁸² *Id.* at 28.

¹⁸³ *Id.* at 180.

¹⁸⁴ *Id.* at 7.

IIFFM issued detailed findings in an interim report (the “2018 IIFFM Report”).

According to the 2018 IIFFM Report, Facebook was an ever-present part of life in Myanmar.¹⁸⁵ It was the “most common social media platform in use in Myanmar”¹⁸⁶ with approximately 20 million users, and became the “main mode of communication.”¹⁸⁷ Information posted on Facebook was further made available through “Facebook Flex,” a data-free service enabling “subscribers to have a text-only version of Facebook.”¹⁸⁸ Thus, Facebook had an immense online presence in Myanmar.¹⁸⁹

The IIFFM noted that Facebook’s platform contained inflammatory “[m]essages portraying Rohingya as violent, dishonest, anti-Bamar, anti-Buddhist, illegal immigrants and/or terrorists.”¹⁹⁰ Additionally, “[d]eath threats, incitement to violence and discrimination, and online harassment are common features . . . [both] against the Rohingya themselves [and] also against moderate commentators, human rights defenders and ordinary people who have views that differ.”¹⁹¹

For example, on October 12, 2016, “Dr. Tun Lwin, a well-known meteorologist with over 1.5 million followers on Facebook, called on the Myanmar people to be united to secure the ‘west gate’ and to be alert ‘now that there is a common enemy.’”¹⁹² Dr. Tun Lwin’s post, referring to the Rohingya, also “stated that Myanmar does not tolerate invaders.”¹⁹³ By August 2018, his post “had 47,000 reactions, over 830 comments and nearly 10,000 shares . . . [as comments also] called for immediate ‘uprooting’ and ‘eradication’ of the Rohingya, citing the situation in Rakhine State as a ‘Muslim invasion.’”¹⁹⁴

In 2018, the IIFFM declared that “[t]he role of social media [during the conflict was] significant. Facebook had been a “useful

¹⁸⁵ *Id.*

¹⁸⁶ *Id.*

¹⁸⁷ *Id.* at 341.

¹⁸⁸ *Id.*

¹⁸⁹ *Id.*

¹⁹⁰ *Id.* at 268, 340.

¹⁹¹ *Id.*

¹⁹² *Id.* at 325.

¹⁹³ *Id.*

¹⁹⁴ *Id.*

instrument for those seeking to spread hate, in a context where, for most users, *Facebook is the Internet* . . . [Additionally,] the response of Facebook has been slow and ineffective.”¹⁹⁵

Though Facebook maintained that its platform was merely an information-sharing vehicle, the 2018 IFFM Report suggested otherwise. Indeed, Facebook’s influence was so strong that many in Myanmar confused Facebook with the Internet itself.¹⁹⁶ Thus, for “many people, Facebook [was] the main, if not only, platform for online news and for using the Internet more broadly,” which made the dissemination of hate speech amenable to Facebook users and their “perception of Facebook as a reliable source of information.”¹⁹⁷

Because of this ubiquity, Facebook was also “a regularly used tool for the Myanmar authorities to reach the public.”¹⁹⁸ Government officials such as the President, State Counsellor, Commander-in-Chief, the Ministry of Information, and the Tatmadaw “rel[ie]d on Facebook to release news and information,” which reinforced the idea that Facebook users could be trusted.¹⁹⁹ The “low digital and social media literacy”²⁰⁰ among the civilian population in Myanmar, in addition to the Government’s reliance on Facebook as a primary mode of communication to share “official announcements,”²⁰¹ led users in Myanmar to believe that Facebook was a source of well-founded information.²⁰²

Though Facebook’s stated goal was to “facilitate[] communication and access to information,” the 2018 IFFM Report found that “the wide reach, relative user anonymity, and difficulty of monitoring or removing posts . . . [made Facebook] a suitable instrument to spread messages that may constitute hate speech.”²⁰³ As a result, it was “unsurprising that propagators of hate speech resort[ed]

¹⁹⁵ *Report of the Detailed Findings of the Independent International Fact-Finding Mission on Myanmar*, *supra* note 171, at 4 (emphasis added).

¹⁹⁶ Paul Mozur, *A Genocide Incited on Facebook, With Posts From Myanmar’s Military*, N.Y. TIMES (Oct. 15, 2018), <https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html>.

¹⁹⁷ *Id.*

¹⁹⁸ *Id.*

¹⁹⁹ *Id.*

²⁰⁰ *Id.*

²⁰¹ *Id.*

²⁰² *Id.*

²⁰³ *Id.*

to Facebook to wage hate campaigns, amplify their message, and reach new audiences.”²⁰⁴ This hate speech advocated “national, racial or religious hatred that constitute[d] incitement to discrimination, hostility or violence.”²⁰⁵ For these reasons, the 2018 IFFM Report concluded that “posts and messages on Facebook have increased discrimination and violence in Myanmar.”²⁰⁶

B. FACEBOOK’S EFFORTS TO CURB USE OF ITS PLATFORM TO SPREAD VIOLENCE WERE SEVERELY LACKING

Crucially, the 2018 IFFM Report concluded that Facebook’s efforts to eliminate hate speech and halt the spread of misinformation in Myanmar fell far short of what was necessary.²⁰⁷

First, the community standards and user agreement policies that Facebook required users to agree to—including rules relating to hate speech and violence—did not actually do anything to halt the use of hate speech on the platform.²⁰⁸ The 2018 IFFM Report suggested that Facebook should further outline specific acts of intervention the company would take in the event the agreed-to policies were violated.²⁰⁹ Moreover, because of these shortcomings, the Report recommended that Facebook implement better data-monitoring systems.²¹⁰

²⁰⁴ *Id.*

²⁰⁵ *Report of the Detailed Findings of the Independent International Fact-Finding Mission on Myanmar*, *supra* note 172 at 331(emphasis added).

²⁰⁶ *Id.* at 342.

²⁰⁷ Indeed, even while conducting its own mission, the IFFM “itself experienced a *slow and ineffective* response from Facebook when it used the standard reporting mechanism to alert the company to a post targeting a human rights defender for his alleged cooperation with the Mission.” *Id.* at 343(emphasis added). The post pertained to a “national traitor,” repeatedly used the term “Muslim,” and was shared and reposted over 1,000 times. *Id.* Comments on the post constituted hate speech as they “explicitly called for the person to be killed, in unequivocal terms.” *Id.* Since Facebook did not take actions, the Mission messaged a Facebook official’s email account; however, the Mission “did not receive a response. Weeks later, Facebook finally took down the reported post, but the Mission “found at least 16 re-posts of the original post still circulating on Facebook.” *Id.* Facebook’s passive efforts to take down flagged content affirms the 2018 IFFM Report’s findings. *Id.* at 341.

²⁰⁸ *Id.* at 342.

²⁰⁹ *Id.* at 341.

²¹⁰ *Id.* at 431-32.

Second, Facebook's method to address "fake accounts and false news" was limited to pre-emptive measures.²¹¹ This resulted in the IFFM report's recommendation that all social media platforms, including Facebook, "should establish early warning systems for emergency escalation, involving relevant stakeholders."²¹² This meant that "[a]ll death threats and threats of harm in Myanmar [should be] treated as serious and immediately removed when detected."²¹³ The IFFM report noted that "early warning systems should be developed and operated transparently and in consultation with key stakeholders, including civil society organizations . . . [and] should be supported by a formal stakeholder group to provide advice and to monitor performance."²¹⁴ According to the IFFM, Facebook and all other social media companies should implement acts of intervention to combat hate speech, as well as prevention.²¹⁵

Third, Facebook was "ineffective [in their] content moderation."²¹⁶ The company was over-reliant on third parties, ill-prepared with a "proper mechanism for emergency escalation, [and displayed] a reticence to engage local stakeholders around systemic solutions and a lack of transparency."²¹⁷ Specifically with respect to Myanmar, Facebook lacked enough content moderators that could interpret and contextually understand local language, and overlooked their "strong . . . unique focus on the Myanmar language and Burman culture."²¹⁸

Fourth, Facebook had failed to "undertake [a] comprehensive human rights impact assessment in Myanmar."²¹⁹ As a result of Facebook's limited efforts, the 2018 IFFM Report recommended that "[a]ll social media platforms active in Myanmar, including messenger systems, should apply international human rights law as a basis for content moderation on their platforms."²²⁰ The UN Guiding Principles

²¹¹ *Id.* at 342.

²¹² *Id.* at 432.

²¹³ *Id.*

²¹⁴ *Id.* at 432.

²¹⁵ *Id.*

²¹⁶ *Id.* at 342.

²¹⁷ *Id.* (citing an open letter to Facebook from six local Myanmar civil society organizations working on the issue of hate speech).

²¹⁸ *Id.* at 343.

²¹⁹ *Id.* at 344.

²²⁰ *Id.* at 431.

on Business and Human Rights “provide[d] a baseline approach” that Facebook should adopt.²²¹

Fifth, country-specific data regarding the spread of hate speech on Facebook’s platform was “imperative to assess the problem and the adequacy of [Facebook’s] response.”²²² However, “Facebook [was] unable to provide” such data, and thus the 2018 IFFM Report also stressed that the company should “develop and implement systems” that could collect and make this information available publicly.²²³

Following the publication of the 2018 IFFM Report, Facebook “publicly acknowledged that the company had been ‘too slow’ in reacting to the concerns raised by civil society organizations”²²⁴ while still refusing to accept any liability beyond its failure to take down posts in a timely manner. Facebook informed the IFFM that “it was trying to solve two specific problems in Myanmar”: hate speech and the spread of false information used for hate speech.²²⁵ Facebook claimed that “technical challenges” prevented it from doing more to address these problems, including “fonts used in Myanmar language,” “improving reporting flows . . . automation . . . and acting on fake accounts” to remove misinformation.²²⁶ As a result, Facebook stated that it had increased the number of local language reviewers and people in the company knowledgeable on Myanmar-specific issues and “put in place a special team working to better understand the specific local challenges and [will] build the right tools.”²²⁷

C. A HUMAN RIGHTS IMPACT ASSESSMENT REPORT
COMMISSIONED BY FACEBOOK DOWNPLAYED ITS
RESPONSIBILITY FOR THE VIOLENCE IN MYANMAR

Following the issuance of the 2018 IFFM Report, Facebook commissioned its own human rights impact assessment report (the

²²¹ *Id.* at 431.

²²² *Id.* at 344.

²²³ *Id.*

²²⁴ *Id.* at 343; see also Alexandra Stevenson, *Facebook Admits It Was Used to Incite Violence in Myanmar*, N.Y. TIMES (Nov. 6, 2018), <https://www.nytimes.com/2018/11/06/technology/myanmar-facebook.html>.

²²⁵ *Report of the Detailed Findings of the Independent International Fact-Finding Mission on Myanmar*, *supra* note 172, at 342.

²²⁶ *Id.*

²²⁷ *Id.*

“HRIA Report”) from the management consulting firm Business for Social Responsibility. The HRIA Report based its methodology on the UN Guiding Principles on Business and Human Rights and identified both Facebook’s actual and potential human rights impacts and how to address them.²²⁸

As the 2018 IFFM Report recommended,²²⁹ the HRIA Report also noted that in response to the Myanmar conflict, “Facebook updated its public-facing Community Standards to include . . . more detail on where the company draws the line on content.”²³⁰ It also “altered its credible violence policies to more proactively delete inaccurate or misleading information created or shared with the purpose of contributing to, or exacerbating, violence or physical harm.”²³¹ The report suggested that as an additional step, Facebook could better implement its Community Standards with more “detailed written guidance about how to enforce the Community Standards in a local context, such as specific slurs, flagged words, and illustrative cases.”²³²

However, while some of its conclusions seemed to comport with those of the 2018 IFFM Report, others appeared designed to absolve Facebook of responsibility for the consequences of its failures in Myanmar and the need to take precautions elsewhere to prevent similar violence. For example, just as the 2018 IFFM Report advocated that Facebook “undertake [a] comprehensive human rights impact assessment in Myanmar,”²³³ the HRIA Report concluded that Facebook should engage in “human rights due diligence” in Myanmar as the country required system-wide change.²³⁴ But the HRIA Report framed its conclusion in the context of a general lack of government accountability and rule of law standards, laws that enabled “systemic

²²⁸ *Human Rights Impact Assessment: Facebook in Myanmar*, BSR 3 (October 2018), https://about.fb.com/wp-content/uploads/2018/11/bsr-facebook-myanmar-hria_final.pdf.

²²⁹ *Report of the Detailed Findings of the Independent International Fact-Finding Mission on Myanmar*, *supra* note 172 at 341, 431-32.

²³⁰ *Human Rights Impact Assessment: Facebook in Myanmar*, *supra* note 228, at 21.

²³¹ *Id.*

²³² *Id.*

²³³ *Report of the Detailed Findings of the Independent International Fact-Finding Mission on Myanmar*, *supra* note 172 at 344.

²³⁴ *Human Rights Impact Assessment: Facebook in Myanmar*, *supra* note 228, at 54.

gender, ethnic, and religious discrimination,”²³⁵ and “deep-rooted and pervasive cultural beliefs in Myanmar . . . [that] reinforce discrimination.”²³⁶

More disturbingly, the HRIA Report also appeared to paint Facebook as a helpless victim of both its own success in becoming the primary means of communication in Myanmar and so-called bad actors who sought to manipulate Facebook’s purportedly neutral communication tools for their own nefarious ends.²³⁷ For example, the report stated that the “implementation of Facebook’s Community Standards present[ed] challenges of a nature and scale never previously addressed by companies or governments.”²³⁸ This was because “with over 2 billion users, this [was] a task of immense complexity and intensity,”²³⁹ and this “challenge [was] even more testing in the Myanmar context”²⁴⁰ where “the majority of the population lack[ed] the digital literacy to effectively navigate the complex world of information-sharing online.”²⁴¹ The HRIA Report thus appeared to insinuate that Facebook could not be held responsible for the fact that its customers in Myanmar were not as experienced with digital tools such as social media as Facebook wanted or expected them to be, or used Facebook’s platform in ways that Facebook had not anticipated—and, even worse, that people in Myanmar simply did not have either the intelligence or sophistication to “correctly” use Facebook—while also seeming to give Facebook a free pass for failing to have had adequate systems in place to fully implement its own community standards.²⁴² It is difficult to read these parts of the HRIA

²³⁵ *Id.* at 18.

²³⁶ *Id.* at 3. According to the HRIA Report, these “deep-rooted and pervasive cultural beliefs in Myanmar . . . reinforce discrimination and . . . result in interfaith and communal conflict.” *Id.* Combined with the fact that there was a “sudden introduction of accessible mobile communication services and [a] rapid growth in access to Facebook,” this “resulted in a steep learning curve for users, policymakers, and civil society.” *Id.* at 24.

²³⁷ *Id.* at 24.

²³⁸ *Id.* at 25.

²³⁹ *Id.*

²⁴⁰ *Id.* at 3.

²⁴¹ *Id.* at 25.

²⁴² *Id.* at 24-25.

Report as anything but classic tech-bro hubris and naivete at its worst.²⁴³

Additionally, while the HRIA Report acknowledged that “Facebook [had become] a useful platform for those seeking to incite violence and cause offline harm,”²⁴⁴ it also minimized this fact by explaining that it was only a “minority of users”²⁴⁵ that sought “to use Facebook as a platform to undermine democracy and incite offline violence, including serious crimes under international law.”²⁴⁶ Specifically, the HRIA Report used the 2018 IFFM Report findings as an example to describe how “Facebook has been used by bad actors to spread anti-Muslim, anti-Rohingya, and anti-activist sentiment.”²⁴⁷

Yet even these weak justifications pale in comparison to what was perhaps the HRIA Report’s most shocking conclusion. Noting that Facebook had increased its number of Myanmar language experts to 60 as of August 2018²⁴⁸— a full year after the acts that prompted the IFFM to call for genocide, crimes against humanity, and war crimes prosecutions with respect to the Rohingya²⁴⁹ — the HRIA Report also cautioned that “[i]nvestment in local staff could raise expectations that Facebook will take a similar approach in other countries.”²⁵⁰ Incredibly, at a time when Facebook “generated \$18.7 billion in revenue, up from \$16.9 billion a year earlier and above analysts’ expectations of \$17.34 billion,”²⁵¹ this warning appears to suggest that Facebook should seriously consider whether implementing the

²⁴³ See Julia Carrie Wong & Matthew Cantor, *How to Speak Silicon Valley: 53 Essential Tech-Bro Terms Explained*, GUARDIAN (June 27, 2019, 1:00 A.M. EDT), <https://www.theguardian.com/us-news/2019/jun/26/how-to-speak-silicon-valley-decoding-tech-bros-from-microdosing-to-privacy>.

²⁴⁴ *Human Rights Impact Assessment: Facebook in Myanmar*, *supra* note 228, at 24.

²⁴⁵ *Id.*

²⁴⁶ *Id.*

²⁴⁷ *Id.*

²⁴⁸ *Id.* at 21.

²⁴⁹ *Id.*

²⁵⁰ *Id.* at 26 (emphasis added).

²⁵¹ Jeff Horwitz, *Facebook Posts Revenue Growth Despite Pandemic*, WALL ST. J. (July 30, 2020), https://www.wsj.com/articles/facebook-fb-2q-earnings-report-2020-11596138406?reflink=desktopwebshare_permalink (reporting on Facebook’s 2020 second-quarter results). Horwitz went on to note that even these astounding figures represented “a *deceleration* from the average gain of nearly 25% for the preceding four quarters” due to the global coronavirus pandemic. *Id.* (emphasis added).

IIFFM’s recommendations in Myanmar—and thereby reducing the potential for further serious human rights abuses—was actually a net positive given that it could result in increased expectations that Facebook would implement those recommendations in other places around the world that ran the risk of descending into violent conflict. Put another way, the HRIA Report seemed to be cautioning Facebook that if it implemented precautionary measures in Myanmar, it would be obligated to apply them globally. That such a conclusion constituted a warning rather than an acknowledgment of the responsibility that comes with having a user base of billions of people—and social media’s singular role in shaping the views and actions of entire societies—is a stark and chilling statement.

D. EVEN AFTER IT WAS ON NOTICE, FACEBOOK CONTINUED TO BE USED FOR HATE SPEECH IN MYANMAR

In 2019, the UN Fact-Finding Mission issued an update and detailed findings to its 2018 report (“2019 IIFFM Report”). The 2019 IIFFM Report found that hate speech directed at ethnic Rakhine had “increased considerably on social media.”²⁵² The Report declared that “Facebook is the leading platform for hate speech in Myanmar.”²⁵³ The 2019 IIFFM Report renewed its call to “Facebook and other social media to enhance their capacity to combat the use of their platforms for the spread and promotion of threats and of hate speech and for the incitement to violence, hostility and discrimination.”²⁵⁴

The 2019 IIFFM Report outlined Facebook’s efforts where it fell short. For example, Facebook removed the pages of 20 individuals and organizations in August 2018 and shut down the official pages of the Arakan Army, the Kachin Independence Army, the Myanmar Democratic Alliance Army, and the Ta’ang National Liberation Army, which Facebook identified as “dangerous organizations.”²⁵⁵ Yet, in a precursor to the ephemeral astroturfing seen in the 2020 Conflict, new

²⁵² Hum. Rts. Council, *Report of the Detailed Findings of the Independent International Fact-Finding Mission on Myanmar*, U.N. Doc. A/HRC/42/CRP.5, at 12 (September 2019).

²⁵³ *Id.*

²⁵⁴ *Id.* at 132.

²⁵⁵ *Id.* at 85-86.

pages with virtual identities quickly replaced those shutdown and they continued to post on Facebook.²⁵⁶

While Facebook responded “positively to removing content amounting to hate speech,” the IFFM declared that “much more is required, especially in preventing and removing hate speech far more quickly and in addressing the spread of removed content that has been reposted prior to removal.”²⁵⁷ The Mission held both the Myanmar government *and* Facebook responsible for “tackling hate speech.”²⁵⁸

VI. CONCLUSION

Social media companies could and certainly should have done more to prevent violence in Artsakh/Nagorno-Karabakh during the 2020 Conflict. Nearly 80 years of international precedent have shown, over and over, that mass media publication and distribution of violent rhetoric aimed at a particular group can provoke hatred and weaponize fear on a widespread scale, resulting in serious violence that amounts to international atrocity crimes. Over that time, international criminal tribunals have held senior media company leaders accountable for what is published on their platforms, and at least one UN-sponsored international fact-finding mission strongly recommended a number of detailed steps it deemed necessary for social media companies to follow in order to mitigate or end the use of their platforms by others to incite violence.

However, it is still the case that international criminal liability is restricted to the relatively narrow circumstances where media executives maintained primary overall editorial control of the publication, allowed the publication of speech that called for the commission of specific acts of violence (as opposed to more general hate speech), conveyed a clear message calling for violence to its intended audience within the particular cultural, linguistic, and political context, and intended to incite genocide or other forms of serious persecution.

Thus, even after 80 years of precedent, warnings, and investigations put all media companies on notice, it is unlikely that

²⁵⁶ *Id.*

²⁵⁷ *Id.*

²⁵⁸ *Id.*

social media companies and their leaders could incur international criminal liability for content posted on their platforms—especially when social media executives maintain public positions of neutrality with respect to posted content, typically maintain some form of content moderation policy that purports to regulate hate speech, violent propaganda, and/or misinformation and disinformation,²⁵⁹ and likely (hopefully) do not share the views of users calling for persecution and execution of groups of people. To date, no international prosecutor has charged any social media executive with international crimes. The repeated failure of social media companies to learn from the past and their own prior mistakes is reckless in the extreme, given the disastrous consequences.

²⁵⁹ Though some social media companies increased measures to combat the hateful and violent content targeting protected groups in response to the criticisms in the IFFM reports and elsewhere, following the primary completion of this Report in 2021, the trend in recent months has been for social media companies to roll back or relax enforcement of their content moderation rules. *See, e.g.,* Anika Collier Navaroli, *I Worked on Twitter’s Rules on Hate Speech. Social Media Platforms are Failing Us Right Now*, BOS. GLOBE (Oct. 18, 2023), <https://www.bostonglobe.com/2023/10/18/opinion/i-worked-twitters-rules-hate-speech-social-media-platforms-are-failing-us-right-now/>; John Herrman, *Why Wartime Social Media Is Hellish and Disorienting*, N.Y.: INTELLIGENCER (Oct. 14, 2023), <https://nymag.com/intelligencer/2023/10/why-wartime-social-media-is-hellish-and-disorienting.html>.