

THE CONSEQUENCES OF GENOCIDE: AN  
INQUIRY INTO INTERNATIONAL CRIMINAL  
LIABILITY OF SOCIAL MEDIA COMPANIES  
FOR ARTSAK

Rajika L. Sh

TABLE OF CONTENTS

- I. INTRODUCTION
- II. BACKGROUND: A HISTORY OF THE
- III. SOCIAL MEDIA AND THE 2020 COVID-19  
A. MANIPULATED SOCIAL MEDIA AND VIOLENT ANTI-  
ARMENIAN RHETORIC  
B. THE RESPONSE FROM SOCIAL MEDIA COMPANIES WAS SLOW AND  
INEFFECTIVE
- IV. INTERNATIONAL CRIMINAL TRIBUNALS AND THE RESPONSIBILITY OF  
COMPANIES AND MASS MEDIA EXECUTIVES  
THEIR ROLE IN ATROCITY CRIMES  
A. THE NUREMBERG IMT FOUND LIABILITY WHERE A  
PUBLISHER CONTINUED TO PUBLISH ARTICLES INCITING  
GENOCIDAL VIOLENCE WHEN A STATE'S  
GENOCIDAL VIOLENCE

4. *The ICTR Appeals Chamber Affirmed Media**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.<sup>3</sup> 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.<sup>4</sup> In Azerbaijan, tens of thousands of people gathered in the capital Baku, demanding the government declare war against

THE CONSEQUENCES OF INACTION: AN INQUIRY

propaganda<sup>13</sup>—

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.<sup>17</sup>

Social media posts stirring up nationalist sentiment simplified the narrative and “contributed to the deepening of hatreds and dehumanization of the other.”<sup>18</sup> 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;<sup>19</sup>

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

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



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.<sup>31</sup> The situation remained in a “simmering stalemate . . . punctuated by armed clashes” until 2020.<sup>32</sup>

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.<sup>33</sup> 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.<sup>34</sup> 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.<sup>35</sup>

### 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>.

<sup>31</sup> See CARLEY, *supra* note 27.

<sup>32</sup> *Nagorno-Karabakh Profile*



“unsupportive of the [Nagorno-Karabakh] peace process. . . .”<sup>36</sup> 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.”<sup>37</sup> Such control was expressed “in three distinct ways: to suppress fundamental human rights, discredit political opponents, and drown out dissenting opinions.”<sup>38</sup> 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.

leaving Twitter mostly unblocked.<sup>42</sup> This prompted a surge of virtual private network (VPN) app downloads in Azerbaijan as citizens tried to circumvent the block.<sup>43</sup> 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.<sup>44</sup>

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.<sup>45</sup> Even online Turkish communities dedicated to sharing content about K-pop music mobilized to spread anti-Armenian hashtags.<sup>46</sup>

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.”<sup>47</sup> The French police had to forcibly disperse the violent

---

<sup>42</sup> 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/>.

<sup>43</sup> 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>.

<sup>44</sup> Pearce, *supra* note 42.

<sup>45</sup> 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.

<sup>46</sup> 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>.

<sup>47</sup> Hume, *supra* note 20.

mob who yelled threats that they were “going to kill the Armenians.”<sup>48</sup> 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.<sup>49</sup>

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.<sup>50</sup> 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.<sup>51</sup>

#### 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

---

<sup>48</sup> *Id.*

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

<sup>50</sup> 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>.

<sup>51</sup> *Id.*

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

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

---

<sup>52</sup> Craig Silverman & Ryan Mac, *It Took Facebook More Than a Year And a Whistleblower*, BUZZF

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.<sup>54</sup> 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.<sup>55</sup> 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.<sup>56</sup>

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—

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].<sup>58</sup>

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.<sup>59</sup> 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.”<sup>60</sup> Twenty-three articles in *Der Stürmer* explicitly called for the “root and branch” extermination of Jewish people,<sup>61</sup> urging that “only when world Jewry had been annihilated would the Jewish problem be solved.”<sup>62</sup>

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*.<sup>68</sup> 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.”<sup>69</sup>

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.”<sup>70</sup> 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.”<sup>71</sup> In acquitting Fritzsche, the court emphasized the significance of the lack of language impelling extermination and knowledge of the atrocities being committed.<sup>72</sup>

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

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,<sup>73</sup> 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.<sup>74</sup> Karl Weinbacher, the second defendant, was Tesch's second-in-command.<sup>75</sup> The third defendant, Joachim Drosihn, was the firm's gassing technician.<sup>76</sup> 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."<sup>77</sup>

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

---

<sup>73</sup> 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).

<sup>74</sup> *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).

<sup>75</sup>





inability to prevent it.<sup>87</sup> Drosihn had a “subordinate position” in the company in relation to his limited knowledge and influence over the “firm’s business activities.”<sup>88</sup> 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.”<sup>89</sup> Ultimately, the Court concluded Drosihn was not in a position at the firm “to influence the transfer of gas to Auschwitz or prevent it.”<sup>90</sup>

C. THE ICTR’S MEDIA CASE HELD MEDIA COMPANIES EXECUTIVES LIABLE FOR INCITING GENOCIDE DUE TO THE MESSAGING D





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.<sup>117</sup>

## 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

“more propitious moment[] to decimate us.”<sup>128</sup> It described the Tutsi as “bloodthirsty” and raised the specter of “Tutsi domination over the Hutu.”<sup>129</sup> 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.”<sup>130</sup> The article then urged Hutus to “become aware of a new Hutu ideology,” “cease feeling pity for the Tutsi,” and follow the ten commandments.<sup>131</sup> Those who did not were explicitly labeled as traitors.<sup>132</sup>

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.<sup>133</sup>

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

] sf a



disclaimers or opposing points of view.<sup>146</sup> 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.”<sup>147</sup> 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.<sup>148</sup>

4.

*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

005190003}TJETQq0.000009





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

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.<sup>161</sup> Where speech was potentially ambiguous in meaning, its “true message” was determined by “how a speech was understood by its intended audience.”<sup>162</sup> If the message remained ambiguous even in context, it could not constitute a direct and public incitement to genocide.<sup>163</sup> It was not necessary that the speech “explicitly call[] for extermination” or be “entirely unambiguous for all types of audiences.”<sup>164</sup>

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.”<sup>165</sup> 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.”<sup>166</sup> 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.<sup>167</sup>

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.”<sup>168</sup> 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.<sup>169</sup>

---

<sup>160</sup> *Id.* at ¶ 695.

<sup>161</sup> *Id.* at ¶¶ 697-98.

<sup>162</sup> *Id.* at ¶¶ 700-01.

<sup>163</sup> *Id.* at ¶ 701.

<sup>164</sup> *Id.* at ¶ 702.

<sup>165</sup> *Id.* at ¶ 706.

<sup>166</sup> *Id.* at ¶ 709.

<sup>167</sup> *Id.*

<sup>168</sup> *Id.* at ¶¶ 757-58.

<sup>169</sup> *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.<sup>170</sup>

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.<sup>171</sup>

Rakhine state in northern Myanmar was historically comprised of two main groups: the Rakhine Buddhists and the Rohingya Muslims.<sup>172</sup> 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."<sup>173</sup> Adding fuel to the fire, Myanmar military soldiers systematically oppressed and persecuted the Rohingya.<sup>174</sup>

---

<sup>170</sup> *Id.* at ¶¶ 345-46.

<sup>171</sup> 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).

<sup>172</sup> 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).

<sup>173</sup> *Id.* at 174.

<sup>174</sup> *Id.*

Following a series of small-scale attacks carried out in August

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.<sup>185</sup> It was the “most common social media platform in use in Myanmar”<sup>186</sup> with approximately 20 million users, and became the “main mode of communication.”<sup>187</sup> Information posted on Facebook was further made available through “Facebook Flex,” a data-free service enabling “subscribers to have a text-

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.”<sup>195</sup>

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.<sup>196</sup> 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.”<sup>197</sup>

Because of this ubiquity, Facebook was also “a regularly used tool for the Myanmar authorities to reach the public.”<sup>198</sup> 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.<sup>199</sup> The “low digital and social media literacy”<sup>200</sup> 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,”<sup>201</sup> led users in Myanmar to believe that Facebook was a source of well-founded information.<sup>202</sup>

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.”<sup>203</sup> As a result, it was “unsurprising that propagators of hate speech resort[ed]

---

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

<sup>196</sup> Paul Mozur,

*Military*, N.Y. TIMES (Oct. 15, 2018),

<https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html>.

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

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.<sup>207</sup>

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.<sup>208</sup> 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.<sup>209</sup> Moreover, because of these shortcomings, the Report recommended that Facebook implement better data-monitoring systems.<sup>210</sup>

---

<sup>204</sup> *Id.*

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

<sup>206</sup> *Id.* at 342.

<sup>207</sup> 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.

<sup>208</sup> *Id.* at 342.

<sup>209</sup> *Id.* at 341.

<sup>210</sup> *Id.* at 431-32.

Second, Facebook's method to address "fake accounts and false news" was limited to pre-emptive measures.<sup>211</sup> 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."<sup>212</sup> This meant that "[a]ll death threats and threats of harm in Myanmar [should be] treated as serious and immediately removed when detected."<sup>213</sup> 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."<sup>214</sup> According to the IFFM, Facebook and all other social media companies should implement acts of intervention to combat hate speech, as well as prevention.<sup>215</sup>

Third, Facebook was "ineffective [in their] content moderation."<sup>216</sup> 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."<sup>217</sup> 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."<sup>218</sup>

Fourth, Facebook had failed to "undertake [a] comprehensive human rights impact assessment in Myanmar."<sup>219</sup> 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."<sup>220</sup> The UN Guiding Principles

---

<sup>211</sup> *Id.* at 342.

<sup>212</sup> *Id.* at 432.

<sup>213</sup> *Id.*

<sup>214</sup> *Id.* at 432.

<sup>215</sup> *Id.*

<sup>216</sup> *Id.* at 342.

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

<sup>218</sup> *Id.* at 343.

<sup>219</sup> *Id.* at 344.

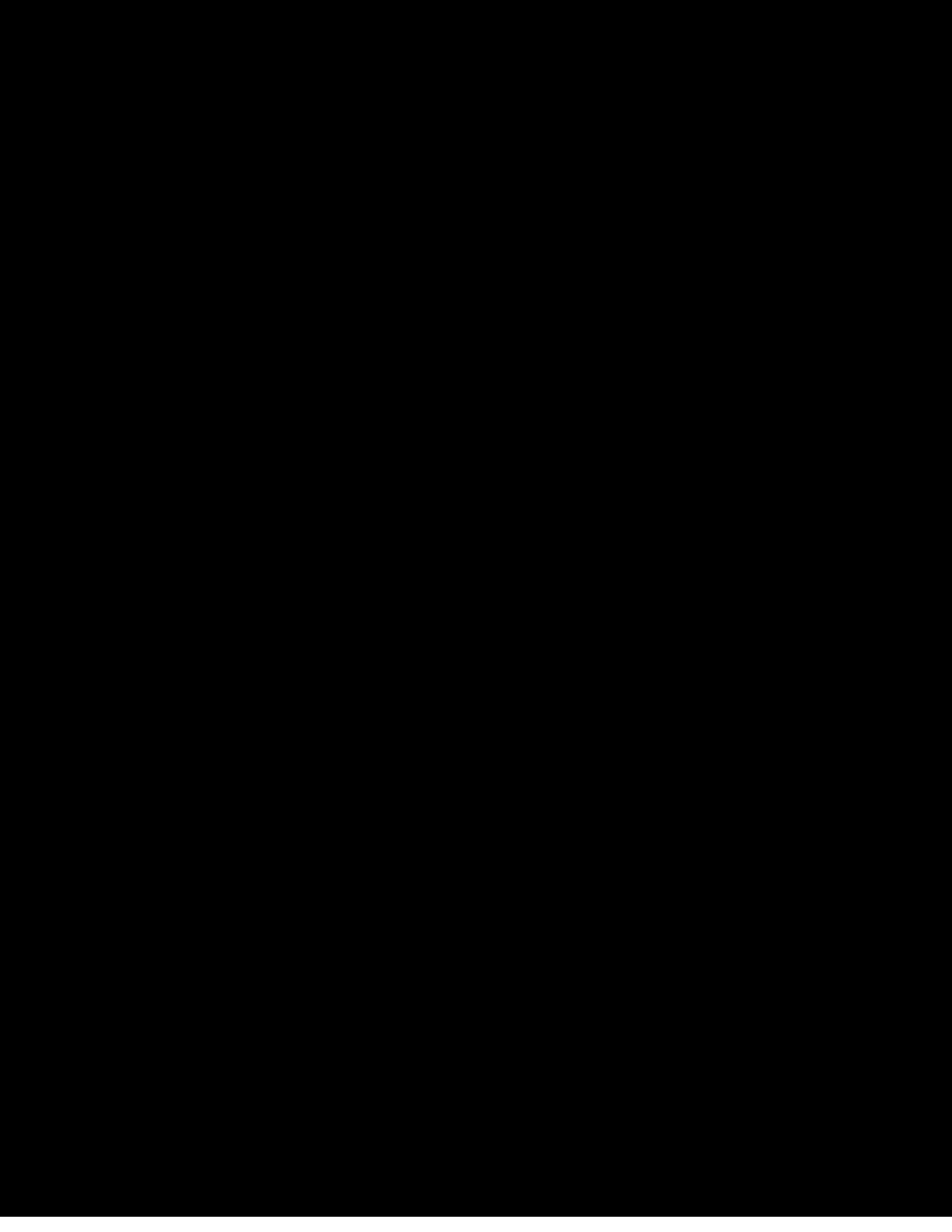
<sup>220</sup> *Id.* at 431.







THE C



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.”<sup>252</sup> The Report declared that “Facebook is the leading platform for hate speech in Myanmar.”<sup>253</sup> 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.”<sup>254</sup>

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.”<sup>255</sup> Yet, in a precursor to the ephemeral astroturfing seen in the 2020 Conflict, new

---

<sup>252</sup> 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).

<sup>253</sup> *Id.*

<sup>254</sup> *Id.* at 132.

<sup>255</sup> *Id.* at 85-86.



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