The Harvard Undergraduate Foreign Policy Initiative

ANALYZING SOCIAL MEDIA MISINFORMATION: INCITEMENT TO ETHNIC VIOLENCE AND HUMAN RIGHTS IMPLICATIONS

Review, case studies, framework, and policies for a rights-based approach to misinformation

Policy Chair: Eleanor Wikstrom
Policy Researchers: Stella Asmerom, Lexa Brenner, April Chen, Maddie Dowd, and Carolyn Wang
Policy Director: Hamaad Mehal

These findings are independent but were presented to Facebook's Misinformation Policy Team.

Spring 2021

Analyzing Social Media Misinformation: Incitement to Ethnic Violence and Human Rights Implications

Review, case studies, framework, and policies for a rights-based approach to misinformation

Harvard Undergraduate Foreign Policy Initiative
Policy Chair: Eleanor Wikstrom
Policy Researchers: Stella Asmerom, Lexa Brenner, April Chen, Maddie Dowd, Carolyn Wang
Policy Director: Hamaad Mehal
These findings are independent but were presented to Facebook's Misinformation Policy Team, April 2021.

Executive Summary

In this research report, our Policy Team aims to provide a workable answer to the following question: In the absence of social media-specific directives from the UN and global governing bodies, how can Facebook use existing human rights frameworks to construct a rights-based approach to misinformation and harm? We begin with a review of existing legislation and human rights guidance with respect to content moderation and misinformation on social media, and proceed to identify gaps in the current ability of international law to combat and measure the consequences of misinformation. We then analyze three international case studies where misinformation on social media has resulted in ethnic violence, and analyze the exact nature of both the types of content shared and the conditions necessary to result in violence and human rights violations. Using the findings of these case studies, we then propose a framework on defining a given post of misinformation as incitement to ethnic violence. Such a framework is then employed to recommend content policies which, in conversation with the current Community Standards, provide a clearer understanding of how Facebook can adopt a human rights-based approach with regards to viral misinformation.

Contents

Policies

Sources

Background

Context

From the Rohingya ethnic cleansing in Myanmar to election-related violence at the United States Capitol, recent events around the globe have highlighted the human rights consequences of viral misinformation on Facebook. Last year, CEO Mark Zuckerberg called for lawmakers to create new rules around harmful content, noting that it was not his role to act as an arbiter of truth or a regulator of free speech. It is apparent that a framework for content moderation with a basis in existing legislation is necessary to protect both the interests of users and the company. Below, we review the ability of the human rights corpus to provide such a foundation for policies which balance concerns of free speech with the need to curb misinformation — defined in this report as an umbrella term to include all false or inaccurate information that is spread on social media, including fake news, rumors, and disinformation[1] — and its real-world consequences.

Review of Existing Legislation and Literature

Existing Concrete Legislation on Content Moderation

When it comes to the regulation of content on social media platforms, we note some smaller multilateral regulations and modifications of domestic legislation. Included below are three key examples:

a. **Anti Fake-News Act (2018) (Malaysia).** The act imposes punishments including fines and jail time against any individual who disseminates "any news, information, data and reports, which is or are wholly or partly false, whether in the form of features, visuals or audio recordings or in any other form capable of suggesting words or ideas."[2]
    i. This legislation is widely criticized by human rights activists and watchdogs, including the United Nations Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. For example, this law does not give defendants an opportunity to prove the truth of their statements and also attempts to hold social media platforms legally liable for content shared through their service, which directly contradicts current international guidance on

[1] Wu, Liang et al. "Misinformation in Social Media: Definition, Manipulation, and Detection." *ACM SIGKDD Explorations Newsletter,* Association for Computing Machinery, Nov 2019. <https://doi.org/10.1145/3373464.3373475>.
[2] "Malaysia: Anti-Fake News Act Comes Into Force." *Global Legal Monitor,* Library of Congress, 19 Apr 2018. <https://www.loc.gov/law/foreign-news/article/malaysia-anti-fake-news-act-comes-into-force/>

propaganda and libel.[3] This approach could lead to even greater violations of human rights through government-backed media censorship.

b. **General Data Protection Regulation (EU).** The GDPR is the largest, and strictest, attempt from an international community to monitor social media by enforcing stricter privacy and security measures; the regulation applies to companies operating in the EU even if the specific users are not citizens of EU countries, and the law operates by imposing fines to companies equal to either 20 million euros or four percent of annual global revenue. Although the GDPR does not address misinformation or harmful speech directly, its ability to limit platforms' access to personal data works such that data privacy laws may render algorithm-driven disinformation "a weapon without a target."[4] In 2019, Facebook was threatened by a $1.6 billion dollar fine due to its security measures.[5] It is worth noting, however, that the GDPR has not culminated in broad reform due to its slow pace and the regulation's current lack of an enforcement mechanism.[6]

c. **Section 230 - Protection for Private Blocking and Screening of Offensive Material (United States).** Section 230 essentially says that companies are not individually responsible for user hate speech or offensive materials, and they therefore cannot be held liable or guilty for user content. This has allowed platforms to host material that is controversial, unpopular, or even outright false — such as in the case of misinformation — without legal liability. Section 230 also outlines general U.S. policy towards free Internet and speech. Some legislators in the United States have pushed to modify Section 230 due to the leniency it gives large social media and tech companies in curbing hate speech and damaging misinformation. Others, however, desire to maintain Section 230 in its current form as the immunity given to tech companies is believed to help business and free speech long-term.[7]

Existing Human Rights / UN Guiding Principles

In the absence of comprehensive national legislation on misinformation and the tech industry, the United Nations has released a few guiding principles on the intersection of global human rights and "fake news," as well as human rights and businesses more broadly.

---

[3] Kaye, David (UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression). "Mandate of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression." *Office of the Special Rapporteur,* United Nations, 3 Apr 2018. <https://www.lawyersforliberty.org/wp-content/uploads/2018/04/OL_MYS_03.04.18-1.pdf>

[4] Campbell, Alex. "How Data Privacy Laws Can Fight Fake News." *Just Security,* Reiss Center on Law and Security at New York University School of Law, 15 Aug 2019. <https://www.justsecurity.org/65795/how-data-privacy-laws-can-fight-fake-news/>.

[5] Koch, Richie. "The GDPR meets its first challenge: Facebook." *GDPR EU,* Proton Technologies AG. <https://gdpr.eu/the-gdpr-meets-its-first-challenge-facebook/>. Accessed 07 Mar 2021.

[6] Coldewey, Devin. "Who Regulates Social Media?" *TechCrunch*, Verizon Media, 19 Oct 2020. <https://social.techcrunch.com/2020/10/19/who-regulates-social-media/>.

[7] Wakabayashi, Daisuke. "Legal Shield for Social Media Is Targeted by Lawmakers." *The New York Times*, 28 May 2020. <https://www.nytimes.com/2020/05/28/business/section-230-internet-speech.html>.

a. **OHCHR Declaration on "Fake News", Disinformation, and Propaganda (2017) [link].** This document, while not legally binding, serves as a primary guiding principle for the United Nations Office of the High Commissioner for Human Rights (OHCHR) when approaching the intersection of media and human rights. The Declaration calls for states to only moderate and control content as explicitly required by international law in order to retain freedom of expression, and states that "intermediaries should never be liable for any content relating to those services" unless they're explicitly commanded to by a court, regulatory body, etc. The Declaration advocates against criminal defamation laws, instead advocating for the imposition of civil penalties where the defendant is given proper legal recourse. This imposes a positive obligation for free speech and media diversity. The Declaration also requires a "clear regulatory framework," such that when companies are proactively regulating content, their policies should be based on "objectively justifiable criteria rather than ideological or political goals and should, where possible, be adopted after consultation with their users."
b. **UN Guiding Principles on Business and Human Rights (2011) [link].** This document is the general recommendation for how businesses should address human rights. The document is essentially a resolution advocating for businesses to respect the human rights corpus and for member States to hold them accountable. In response to this publication, Facebook recently launched a corporate human rights policy which "sets out the human rights standards we will strive to respect as defined in international law" and includes establishing an independent Oversight Board, publicly sharing Facebook's human rights work, and changing key content policies, among other things.[8]

Criticisms of the Gaps in the Human Rights Corpus

While the human rights corpus does provide guiding principles on the intersection of media, business, and human rights, there are not any broad international treaties or agreements that explicitly govern misinformation. A 2019 literature review from the Human Rights, Big Data and Technology project noted that "very little research" is available on the human rights impact of misinformation, and that responses to this phenomena by states and companies "are lacking a human rights approach and sometimes they can actually pose further threats to human rights."[9] In the place of a codified set of content guidelines and enforcement frameworks, countries often rely on social media and tech companies to self-govern. However, this leads to a lack of accountability and external monitoring, and also risks imposing the values of a small group of employees on half the world's population.[10]

---

[8] Sissons, Miranda. "Our Commitment to Human Rights." *Facebook Newsroom,* Facebook, 16 Mar 2021. <https://about.fb.com/news/2021/03/our-commitment-to-human-rights/>.
[9] "Addressing the Human Rights Impact of Misinformation Without Further Harming Human Rights." *HRBDT,* The Human Rights, Big Data and Technology Project. <https://www.hrbdt.ac.uk/addressing-the-human-rights-impacts-of-misinformation/> Accessed 23 Feb 2021.
[10] Benesch, Susan. "But Facebook's Not a Country: How to Interpret Human Rights Law for Social Media Companies." *Yale Journal on Regulation,* 14 September 2020.

Even if there was a comprehensive set of legislation, the international community is limited in its ability to respond to human rights abuses in a timely manner. Compliance with the International Covenant on Civil and Political Rights (a major element of the International Bill of Human Rights) is monitored by the UN Human Rights Committee; however, because violations often rely on countries self-reporting themselves, the Committee is limited in its knowledge of human rights abuses. Additionally, although the Committee is able to publish their findings and give recommendations, they are unable to enforce their rulings or impose punishment against actors who violate the ICCPR.[11]

Clearly, there is a need for human rights-based guidance on misinformation that is also actionable and enforceable by companies themselves. Recognizing this, human rights and misinformation experts like Susan Benesch of the Dangerous Speech project have proposed imposing a modified, standard version of content moderations for tech companies that relies on codified articles of the human rights corpus. While most of the proposal is concerned with dangerous speech more broadly, limiting access to misinformation in the interest of public health (i.e. in the current Covid-19 pandemic) is regarded as a legitimate reason to restrict speech under international law, particularly ICCPR Article 19(3)(b). Theoretically, a human rights framework would increase transparency between companies and governments, as well as lead to better enforcement of nebulous regulations without unfairly imposing the values of a small group of private employees on half the world's population.[12]

Overview of Related Human Rights Articles

In formulating our own human rights-based approach to content moderation and misinformation policy, we begin by providing an overview of the relevant articles in the human rights corpus.

International Bill of Human Rights

The International Covenant for Civil and Political Rights (ICCPR) and the International Covenant for Economic, Social, and Cultural Rights are legal treaties created to underpin the Universal Declaration of Human Rights (UDHR). Together, the three documents compose the International Bill of Human Rights, widely considered to be a major source of international law. The ICCPR generally deals with primary rights, often negative, such as the right to free speech or life. Because the ICCPR is concerned more with what a government *cannot* do to its citizens rather than the privileges it bestows, it is the most relevant core document for the relationship

<https://www.yalejreg.com/bulletin/but-facebooks-not-a-country-how-to-interpret-human-rights-law-for-social-media-companies/>.

[11] Persaud, Santhosh. "Protecting Refugees and Asylum Seekers under the International Covenant on Civil and Political Rights."*New Issues in Refugee Research,* UNHCR, Research Paper No. 132, Nov 2006. <https://www.unhcr.org/research/working/4552f0d82/protecting-refugees-asylum-seekers-under-international-covenant-civil-political.html>.

[12] Benesch, "Facebook," 2020.

between media companies, content moderation, and human rights.[13] Here are the articles which experts like Susan Benesch have identified[14] as most important in the discussion of online content moderation:

a. ICCPR Article 19[15]

1. Everyone shall have the right to hold opinions without interference.
2. Everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice.
3. The exercise of the rights provided for in paragraph 2 of this article carries with it special duties and responsibilities. It may therefore be subject to certain restrictions, but these shall only be such as are provided by law and are necessary:
    (a) For respect of the rights or reputations of others;
    (b) For the protection of national security or of public order (ordre public), or of public health or morals.

b. ICCPR Article 20[16]

1. Any propaganda for war shall be prohibited by law.
2. Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law.

For part 2 of Article 20, there has been considerable debate about the severity of hate speech necessary before a state can intervene. According to the Committee, incitement is a higher threshold than just derogatory or offensive speech, and states can only preclude content that leads to such incitement under Article 20.[17]

By demonstrating concern for both freedom of speech and hate speech, these two articles are most formative in framing the discussion about human rights and online content moderation. However, they have also been criticized for their nebulous language and lack of enforceability in the current format, indicating a need for further development into a clear and standardized rights-based framework of online content moderation.[18] These two articles are particularly relevant for our study of misinformation because of its potential to incite public violence, a violation of ICCPR Article 19(3)(b); its potential to incite direct hate speech, a violation of ICCPR Article 20(2); and its potential to incite discrimination which infringes upon the rights of others, a violation of ICCPR Article 19(3)(a). The last point is discussed in more detail in the following section. The bulk of this report is dedicated to case studies which demonstrate the ability of misinformation, which is problematic but not necessarily a human rights infringement by itself, to lead to violations of international law as outlined above. **By re-framing our**

[13] Ibid.
[14] Ibid.
[15] United Nations General Assembly. *International Covenant on Civil and Political Rights*, 19 December 1966. <https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx>.
[16] Ibid.
[17] Global Campaign for Free Expression. (n.d.). *Towards an interpretation of article 20 of the ICCPR: Thresholds for the prohibition of incitement to hatred Work in Progress*. The regional expert meeting on article 20, Organized by the Office of the High Commissioner for Human Rights,.
<https://www.ohchr.org/Documents/Issues/Expression/ICCPR/Vienna/CRP7Callamard.pdf>
[18] Benesch, "Facebook," 2020.

understanding of misinformation as a tool with the potential to infringe on a variety of human rights, and identifying which aspects of said misinformation make such violations possible, we hope to contribute to a more robust framework of content moderation that is rooted in the desire to protect international human rights.

Human Rights Articles and Misinformation

Beyond its connection to Articles 19 and 20 of the ICCPR, and extending ICCPR Article 19(3)(b), mis/disinformation on social media is a human rights issue because it can lead to infringement on a variety of rights that are explicitly outlined in the International Bill of Human Rights and other international treaties. Groups like Global Partners Digital[19] have established some of the linkages between outcomes of misinformation and particular articles in the human rights corpus, including the right to free and fair elections (Article 25, ICCPR), as seen during the 2016 and 2020 elections in the United States; the right to health (Article 12, ICESCR), as seen during the current coronavirus pandemic; the right to freedom from unlawful attacks upon one's honour and reputation (Article 17, ICCPR), as seen in disinformation which directly threatens journalists; and the right to non-discrimination (Article 26, ICCPR), as seen in disinformation which targets specific groups in society and is designed to incite violence. **In our report, we will focus on the final element of this list: the relationship between misinformation which targets particular ethnic groups and outcomes of real-world violence and harm.**

Human Rights Articles and Ethnic Violence

In focusing primarily on the subset of misinformation which can be directly or indirectly linked to ethnic violence, we first review relevant international law and human rights articles in that subject area.

Although ethnic violence — and its extreme form, ethnic cleansing — are not officially defined nor denoted as specific crimes under the ICCPR or other treaties, the acts associated with ethnic violence do violate international laws. For ethnic violence more broadly, several provisions of the ICCPR in particular prohibit discrimination based on minority status, including Article 26 and Article 27:

> **Article 26[20]**
> All persons are equal before the law and are entitled without any discrimination to the equal protection of the law. In this respect, the law shall prohibit any discrimination and guarantee to all persons equal and effective protection against discrimination on any ground such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status.

[19] Wingfield, Richard. "A Human Rights Approach to Disinformation." *Global Partners Digital,* 15 Oct 2019. <https://www.gp-digital.org/a-human-rights-based-approach-to-disinformation/>.
[20] United Nations General Assembly. *International Covenant on Civil and Political Rights*, 19 December 1966. <https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx>.

> **Article 27**[21]
> In those States in which ethnic, religious or linguistic minorities exist, persons belonging to such minorities shall not be denied the right, in community with the other members of their group, to enjoy their own culture, to profess and practise their own religion, or to use their own language.

In one example of the intersection between content moderation and human rights, the UN Special Rapporteur on Minority Issue, Fernand de Varennes, proposed that tech companies — such as Facebook — should police hate speech related to minorities more thoroughly. According to Varennes, failing to do so is a potential violation of Article 27.[22]

Beyond general ethnic violence, ethnic cleansing in particular is often associated with infringement on treaties and articles related to genocide, property damage, and sexual violence.[23] With regards to genocide, the International Criminal Court (ICC) has authority to investigate genocide accusations, and is limited to the definition of genocide as given in Article II of the Convention on the Prevention and Punishment of the Crime of Genocide, included below.

> **Article II**[24]
> In the present Convention, genocide means any of the following acts committed with intent to destroy, in whole or in part, a national, ethnical, racial or religious group, as such:
> (a) Killing members of the group;
> (b) Causing serious bodily or mental harm to members of the group;
> (c) Deliberately inflicting on the group conditions of life calculated to bring about its physical destruction in whole or in part;
> (d) Imposing measures intended to prevent births within the group;
> (e) Forcibly transferring children of the group to another group.

An important section of Article 2 is the use of the word "intent." When prosecuting states for genocide, international courts must prove that the accused government acted with a specific intent or mindset to commit genocide, which is a relatively high standard. Susan Benesch has proposed a framework[25] to define such incitement to genocide for a prosecutorial purpose in a clear and actionable manner while still balancing the concerns of free speech. Our own final

[21] Ibid.
[22] United Nations Newsroom. "Hate speech on Facebook poses 'acute challenges to human dignity' — UN expert." *UN News Human Rights,* United Nations, 23 December 2020. <https://news.un.org/en/story/2020/12/1080832>.
[23] "Ethnic Cleansing." UN Office on Genocide Prevention and the Responsibility to Protect, United Nations. <https://www.un.org/en/genocideprevention/ethnic-cleansing.shtml>. Accessed 07 Mar 2021.
[24] United Nations General Assembly. *Convention on the Prevention and Punishment of the Crime of Genocide,* 9 December 1948. <https://www.un.org/en/genocideprevention/documents/atrocity-crimes/Doc.1_Convention%20on%20the%20Prevention%20and%20Punishment%20of%20the%20Crime%20of%20Genocide.pdf>
[25] Benesch, Susan. "Vile Crime or Inalienable Right: Defining Incitement to Genocide." *Virginia Journal of International Law, 48(3),* 2008.

framework, modified to address content moderation by social media platforms, is based upon her principles to recognize dangerous speech.

Case Studies

Notes on Ethnic Violence and the Global South

Recognizing Facebook's advances in combatting health-related misinformation, we turn our attention to proposing a human rights-based approach to misinformation that potentially incites ethnic conflict. In doing so, we will investigate the role of misinformation on social media in making the issue of ethnic tension salient through three primary case studies: the Rohingya people of Myanmar, the Fulani people of Nigeria, and Hindu nationalism in India. Prior to these case studies, however, it is important for us to acknowledge our positionality as American university students researching ethnic violence in three Global South countries. Our goal is not to problematize the Global South by suggesting that ethnic conflict is exclusive to those regions; indeed, issues of ethnicity and race are alive and well in the United States too, evidenced in nationwide protests during Summer 2020 against anti-Black racism and police brutality as well as the rise in anti-Asian hate crimes due to racialized rhetoric and misinformation regarding the coronavirus. Rather, we seek to offer three examples in regions where Facebook's user base is growing most rapidly and where the platform plays a large role in infrastructure (for example, the Free Basics program in Myanmar which renders Facebook virtually synonymous with the internet).

Furthermore, we wish to refute the primordialist interpretation of ethnicity which considers ethnic conflict to be innate, inevitable, and even primitive. Instead, we employ a constructivist lens which acknowledges that ethnicity and ethnic conflict are often creations of the state and political elites. We do so to avoid the interpretation that outbreaks of ethnic violence via misinformation are a result of "backwardness." Rather, it is important to keep in mind that political, economic, and environmental conditions throughout history have informed the creation of ethnic divides which are then amplified by the simultaneously connective (i.e. highlighting differences) and isolating (i.e. creating echo chambers) nature of social media. This aligns with observations by scholars of ethnic politics, who have noted that "with substantial increases in communication and intergroup contact [via media technology], the divisive sense of ethnonational uniqueness has been reinforced rather than dissipated."[26]

[26] Jalali, Rita, and Seymour Martin Lipset. 1992. "Racial and Ethnic Conflicts: A Global Perspective." Political Science Quarterly 107(4): 595.

Myanmar

Background on the Rohingya Conflict

The Rohingya people are a heavily marginalized Indo-Aryan, Muslim minority group in Myanmar, almost all of whom have traditionally inhabited the western state of Rakhine, also known as Arakan. Very little about their history has not been politicized and thoroughly obfuscated: academic sources are unclear about much of their past, while non-academics almost uniformly advocate some theory aimed at advancing a political goal. The Arakan Rohingya National Organization,[27] a pro-Rohingya activist group, claims that the Rohingya have lived in Arakan "since time immemorial," while most academics agree that the first centralized government in Rakhine state was of ethnic Indo-Aryans, and that the first ethnic Burmese to arrive there did so in about the tenth century.[28] The time and method of the Rohingya's conversion to Islam, meanwhile, is an issue about which there is almost no scholarly consensus at all.

Contrary to this view and against most academic historical evidence, Burmese nationalists contend that the Rohingya are in fact migrants from Bengal, who were allegedly brought to what is now Myanmar as laborers by the British in the nineteenth century. While this overall narrative is discredited — there is evidence of Rohingya presence in Rakhine well before the nineteenth century — the issue of British migrant labor is not fictitious, and indeed plays a large role in present-day ethnic tensions within Myanmar. Nineteenth-century Myanmar was administered as a part of British India, and indentured laborers were taken from Bengal to Myanmar, provoking a nationalist backlash, and the animosity from that time lingers to this day.[29]

When Myanmar – then known as Burma – became independent of British rule in 1948, the Rohingya were able to participate in the political life of the country, a right guaranteed by Article 21 of the UDHR (signed by Myanmar in the same year).[30] The area they lived, Rakhine, even achieved statehood in 1974. In 1977, however, the government launched Operation Dragon King (Naga Min) in Rakhine State, justifying the operation by considering the Rohingya "illegal immigrants."[31] The president of the Burmese Rohingya Organization UK says that the violence against the Rohingya is driven by the fact that the Rohingya "are a different ethnic group. They

---

[27] Arakan Rohingya National Organization <https://www.rohingya.org/>.
[28] Al Jazeera Staff. "Who are the Rohingya?" *Al Jazeera,* Al Jazeera Media Network, 18 Apr 2018. <https://www.aljazeera.com/features/2018/4/18/who-are-the-rohingya>.
[29] Ibid.
[30] "International Law." *Burma Link,* 27 Oct 2014. <https://www.burmalink.org/background/burma/international-crimes-and-impunity/international-law/#:~:text=While%20not%20a%20treaty%2C%20UDHR,Declaration%20on%20December%2010%2C%201948>.
[31] "Timeline: A visual history of the Rohingya refugee crisis." *Doctors Without Borders News & Stories,* Medecins Sans Frontieres / Doctors Without Borders, 21 Aug 2020. <https://www.doctorswithoutborders.org/what-we-do/news-stories/news/timeline-visual-history- rohingya-refugee-crisis>.

have a different appearance and religion."[32] Operation Dragon King drove around 200,000 Rohingya into Bangladesh in a one-year period, and Bangladesh set up refugee camps where the Rohingya could receive medical aid.[33]

Eventually, Bangladesh began to repatriate the refugees and most were returned to Myanmar, where, in 1982, the military regime stripped them of their citizenships and ethnic minority status under legislation commonly known as the 1982 Citizenship Law.[34] After a military crackdown in 1989 (during which Burma was renamed "Myanmar"), the military increased its presence in northern Rakhine state, and the Rohingya begin to experience further human rights violations in the form of compulsory labor, forced relocation, rape, executions, and torture.[35] In 1991, the military launched Operation Clean and Beautiful Nation (Pyi Thaya),[36] which officially authorized soldiers to commit widespread violence. About 250,000 Rohingya fled to Bangladesh. During this same time period, the government also created a special border security force called NaKaSa to harass and persecute Rohingya seeking refuge in Bangladesh.

Up until 1992, refugees continued to arrive in Bangladesh. That year, however, Myanmar and Bangladesh signed a repatriation agreement, and the country closed to new arrivals. Over the following years, hundreds of thousands of Rohingya were sent back to Myanmar. Meanwhile, in Myanmar, the situation became worse, with the government refusing to issue birth certificates to babies with Rohingya parents in 1994. In 1995, the government began to issue a form of identification for Rohingyas known as a temporary registration card or "white card." It did not serve as proof of citizenship.[37] Persecution persisted for the following decades, and even after the election of Nobel Peace Prize laureate Aung San Suu Kyi, the Rohingya were excluded from the census in 2014. In 2012, there was a surge of violence against the minority after the Rakhine Nationalities Development Party and All-Arakanese Monks' Solidarity Conference issued statements urging townships to band together to solve the Rohingya "problem" and even offering explicit orders for citizens to secure Rohingya villages, resulting in further human rights violations including "arbitrary detention, forced labor, rape, torture, forcible relocations," among others.[38]

---

[32] Persio, Sofia Lotto. "Why Myanmar Hates the Rohingya." *Newsweek World,* Newsweek Digital, 15 Sept 2017. <https://www.newsweek.com/why-myanmar-hates-rohingya-665883>.
[33] Doctors Without Borders, "Timeline," 2020.
[34] Burmese Rohingya Organization UK. "Myanmar's 1982 Citizenship Law and Rohingya." *Burma Campaign,* Dec 2014. <https://www.burmacampaign.org.uk/media/Myanmar%E2%80%99s-1982-Citizenship-Law-and-Rohingya.pdf>.
[35] Doctors Without Borders, "Timeline," 2020.
[36] United States Holocaust Memorial Museum. "Timeline." *Burma's Path to Genocide,* USHMM. <https://exhibitions.ushmm.org/burmas-path-to-genocide/timeline>. Accessed 03 Mar 2021.
[37] Ibid.
[38] Human Rights Watch. "'All You Can Do is Pray: Crimes Against Humanity and Ethnic Cleansing of Rohingya Muslims in Burma's Arakan State." *Human Rights Watch,* 22 Apr 2013. <https://www.hrw.org/report/2013/04/22/all-you-can-do-pray/crimes-against-humanity-and-ethnic-cleansing-rohingya-muslims>.

In 2014, ultra-nationalist monks formed MaBaTha,[39] a social and religious movement aimed at preserving Buddhist values. This movement became the most prominent anti-Muslim and anti-Rohingya organization in Myanmar. In 2015, the Rohingya's white cards were invalidated and they were forced to register for new national verification cards that identify them as immigrants from Bangladesh.

Overview of the Rohingya Ethnic Cleansing / Genocide

In 2016, the situation of the Rohingya escalated to a situation which the United Nations termed a "textbook example of ethnic cleansing."[40] After a small group of Rohingya men attacked a police post in Myanmar in October 2016, the military launched a violent crackdown in the northern Rakhine state which involved burning Rohingya villages and attacking civilians. Similar military operations continued throughout 2017, leading over 700,000 Rohingya to flee Myanmar and over 120,000 more to become internally displaced. A 2018 study estimated that the military and local population killed at least 25,000 Rohingya and perpetrated sexual violence against 18,000 Rohingya women and girls.[41] The persecution of the Rohingya at the hands of the Myanmar military has since been the subject of a genocide case levied against Myanmar at the International Court of Justice.[42]

[39] International Crisis Group. "Report No. 290: Buddhism and State Power in Myanmar." *International Crisis Group,* 5 Sept 2017. <https://www.crisisgroup.org/asia/south-east-asia/myanmar/290-buddhism-and-state-power-myanmar>

[40] United Nations Newsroom. "UN human rights chief points to 'textbook example of ethnic cleansing' in Myanmar." *UN News: Migrants and Refugees,* United Nations, 11 Sept 2017. <https://news.un.org/en/story/2017/09/564622-un-human-rights-chief-points-textbook-example-ethnic-cleansing-myanmar>.

[41] Habib, Mohshin. and Jubb, Christine. and Ahmad, Salahuddin. and Rahman, Masudur. and Pallard, Henri. and Ontario International Development Agency, issuing body. *Forced migration of Rohingya : the untold experience / Mohshin Habib, Christine Jubb, Salahuddin Ahmad, Masudur Rahman, Henri Pallard ; photographs, Salahuddin Ahmad* Ontario International Development Agency, Canada Ottawa, Ontario 2018.

[42] Human Rights Watch. "Questions and Answers on Gambia's Genocide Case Against Myanmar before the International Court of Justice." Human Rights Watch, 5 Dec 2019. <https://www.hrw.org/news/2019/12/05/questions-and-answers-gambias-genocide-case-against-myanmar-international-court#>.

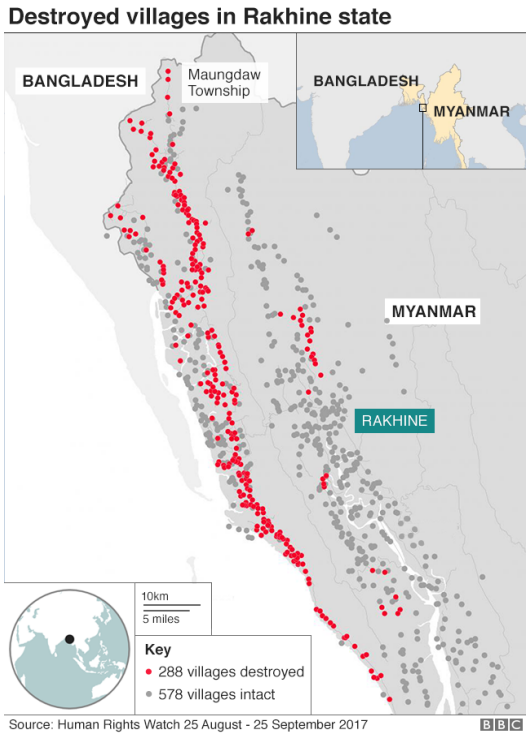


Diagram: Map of destroyed villages in the Rakhine state, where most Rohingya were concentrated during the ethnic cleansing. Source: Human Rights Watch

Linking Misinformation to the Conflict

In June of 2016, Facebook launched Facebook Flex and Free Basics in Myanmar. These services offered access to basic online services without online data charges, essentially rendering Facebook synonymous to the Internet in Myanmar. Three months later, in October, the attack of police posts by Rohingya incited a wave of violence against Rohingya communities, causing about 86,000 Rohingya to flee the country. In the wake of this violence, anti-Muslim rhetoric on Facebook escalated, although mostly in the form of personal attacks rather than attacks on the Rohingya community. A year after the 2016 attacks, another group of Rohingya attacked more police posts, provoking an even stronger crackdown by the military which sent hundreds of thousands of Rohingya across the border into Bangladesh. This second crackdown was accompanied by a series of anti-Rohingya Facebook posts from beauty queens, faux government groups, real government groups, and news organizations for the purpose of fostering public animosity against the Rohingya by labeling them "terrorists" and spreading rumors about their alleged acts of violence. An investigation by the New York Times[43] details how military officials posed as ordinary citizens on Facebook, using fake accounts to spam venomous comments and share misinformation posted by more popular accounts to increase engagement and help such posts spread rapidly. Even after the situation was termed an ethnic cleansing by the United

[43] Mozur, Paul. "A Genocide Incited on Facebook, with Posts from Myanmar's Military." *The New York Times,* 15 Oct 2018. <https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html>.

Nations, incendiary misinformation continued to go unchecked on Facebook. The quick facebook search "rohingya terrorist" in 2021 still reveals several videos from the height of the period of ethnic cleansing which have not been removed or flagged with a content warning despite the opening of a genocide case against Myanmar at the International Court of Justice. Facebook has since adjusted its policies in Myanmar to mitigate direct hate speech,[44] but misinformation in the form of misattributed graphic videos remains unchecked on the platform.

Because this period of ethnic cleansing and its relation to Facebook are well-defined, we provide a timeline of events below. All sources for dates and events mentioned in the previous paragraph are included in the timeline along with examples of Facebook posts, many found by our own HUFPI research team, to illustrate the environment on Facebook during that time.

Timeline

June 7, 2016 - Facebook partners with MPT,[45] a state-run, leading telecommunications company in Myanmar, to launch Free Basics and Facebook Flex in Myanmar.

June 23, 2016 - a Buddhist mob in Bago reportedly destroys homes and forces dozens of villagers to flee[46] after rumors spread on Facebook that a new building in the village was going to be a Muslim school.

Oct. 14, 2016 - A group of Rohingya men attack a number of police posts, killing nine officers. In response, the Myanmar Military launches a violent "clearance operation" in Northern Rakhine. They kill people, rape women, and burn down villages. About 86,000 Rohingya flee to Bangladesh.[47]

Mar. 2017 - A Facebook post accuses a Muslim man, Harry Myo Lin, of seducing a Buddhist girl. Afterwards, he began to receive death threats.[48]

Aug. 25, 2017 - Muslim insurgents now known as the Arakan Rohingya Salvation Army (ARSA) attack 30 police posts and an army base in the north of Rakhine State, killing 12

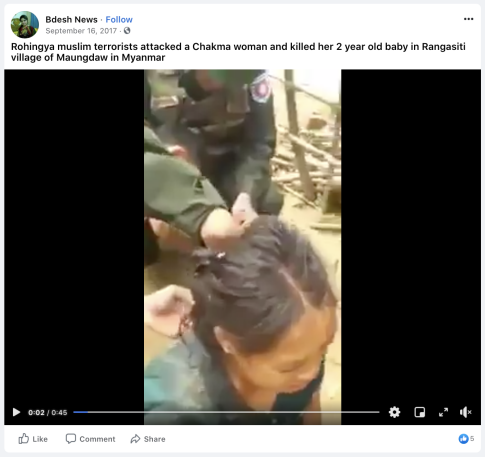
[44] Frankel, Rafael (Facebook Director of Policy, APAC Emerging Countries). "An Update on the Situation in Myanmar." *Facebook Newsroom,* Facebook, 11 Feb 2021. <https://about.fb.com/news/2021/02/an-update-on-myanmar/>.
[45] Waring, Joseph. "Myanmar's MPT Launches Facebook's Free Basics." *Mobile World Live*, 7 June 2016, <www.mobileworldlive.com/asia/asia-news/myanmars-mpt-launches-facebooks-free-basics.>
[46] Mon, Ye. "Residents Flee, police move in after mob violence targets Muslims in Bago village." *Myanmar Times,* The Myanmar Times, 27 Jun 2016. <http://www.mmtimes.com/index.php/national-news/21055-residents-flee-police-move-in-after-mob-violence-targets-muslims-in-bago-village.html>.
[47] USHMM, "Timeline," 2021.
[48] Rajgopalan, Megha. "Internet Trolls Are using Facebook To Target Myanmar's Muslims." *BuzzFeed News,* BuzzFeed, 18 Mar 2017. <https://www.buzzfeednews.com/article/meghara/how-fake-news-and-online-hate-are-making-life-hell-for#.wlGyPB4gk>.

officers. In response, the Myanmar Military, "backed by local Buddhist mobs," began to burn villages and attack civilians.[49]

Sept. 2, 2017 - The government says that more than 2,600 houses were razed in Rohingya-majority areas in the week following Aug. 25.[50]

Sept. 11, 2017 - The U.N. High Commissioner for Human Rights calls the military operation "a textbook example of ethnic cleansing"[51] and declares the Rohingya crisis to be an emergency due to the high number of refugees fleeing to Bangladesh.

Sept. 16, 2017 - This video (screenshot to the right) is posted as evidence that "Rohingya terrorists" are committing violence against local Buddhist citizens. The soldiers in the video seem to actually be Myanmar Military, not Rohingya (see image of Myanmar Military Crest in this article). The HUFPI research team found and reported this post on March 23, 2021, and it has since been taken down, so the link doesn't work anymore. The account that posted it, however, is still up.



Sept. 18, 2017 - This still-available Facebook post declares that 100,000 "Rohingya terrorists" are a major threat to Modi, indicating the permeation of anti-Rohingya sentiment outside of Myanmar's borders.

Sept. 19, 2017 - Myanmar leader Aung San Suu Kyi vows to punish the perpetrators of rights violations but does not address U.N. accusations of ethnic cleansing. By now, about 410,000 Rohingya have fled to Bangladesh.[52] This still-available Facebook video (content warning: serious gore) by beauty queen Shwe Eain Si posted around this time shows that celebrities were perpetuating the state-sponsored narrative that the ARSA was killing mass numbers of Buddist citizens and needed to be stopped.

Oct. 9, 2017 - This still-available Facebook post reports on Hindus escaping "death squads of Rohingya terrorists" in Myanmar and being housed in separate refugee camps in Bangladesh.

---

[49] BBC News Staff. "Myanmar: What sparked latest violence in Rakhine?" *BBC News Asia,* BBC, 19 Sept 2017. <https://www.bbc.com/news/world-asia-41082689>
[50] Reuters Staff. "Timeline: Three years on, a look at the Rohingya crisis." *Reuters,* 20 Aug 2020. <https://www.reuters.com/article/us-myanmar-rohingya-timeline/timeline-three-years-on-a-look-at-the-rohingya-crisis-idUSKBN25H03Y>
[51] UN Newsroom, "UN human rights chief," 2017.
[52] BBC News, "Myanmar," 2017.

Oct. 10, 2017 - This still-available Facebook post from India Today reports on "Rohingya Terror Caught on Camera."

Oct. 12, 2017 - Army commander-in-chief, Senior General Min Aung Hlaing, says the Rohingya Muslims are not natives of Myanmar during a meeting with U.S. ambassador Scot Marciel.[53]

Oct. 22, 2017 - This still-available Facebook post claims to document the Rohingya on their way to terrorize a Myanmar village. On the same day, the same account posts another video linking the Rohingya to ISIS. By now, about 200,000 more refugees have fled to Bangladesh.[54]

Nov. 8, 2017 - This still-available Facebook post shows a video from inside a "Jihadist Terrorist" camp for Rohingyas.

Nov. 16, 2017 - This still-available Facebook post shows a woman testifying about her village being attacked by "Rohingya Muslim Terrorists."

Dec. 21, 2017 - The United States imposes sanctions on 13 "serious human rights abusers and corrupt actors," including the general who oversaw the crackdown against the Rohingya.[55]

Nov. 5, 2018 - Facebook admits it was used to "foment division and incite offline violence" in Myanmar.[56]

Summary of Trends

During the 2016-17 ethnic cleansing, misinformation on Facebook amplified the potential for violence against the Rohingya in Myanmar, both state-sponsored and mob-driven. Misinformation in the form of misattributed graphic videos, shared by influential public figures and spammed by fake accounts, contributed to the fostering of widespread violent sentiment and justified military attacks against an already marginalized population.

[53] Reuters, "Timeline," 2020.
[54] USA for UNHCR Staff. "Rohingya Refugee Crisis Timeline." *USA for UNHCR,* United Nations Refugee Agency, 25 Aug 2019. <https://www.unrefugees.org/news/rohingya-refugee-crisis-timeline/>.
[55] Reuters, "Timeline," 2020.
[56] Warofka, Alex. "An Independent Assessment of the Human Rights Impact of Facebook in Myanmar." *Facebook Newsroom,* Facebook, 5 Nov 2018. <https://about.fb.com/news/2018/11/myanmar-hria/>.

Nigeria

Background on the Fulani Conflict

Nigeria is Africa's most populous country with over 200 million inhabitants and the largest Sub-Saharan national economy.[57] It is in many ways an artificial state. The geographical area that currently composes Nigeria was formed after the amalgamation of two British colonies — Northern and Southern Nigeria — in 1914. The combined territories became an independent state in 1960.[58] However, unifying borders in no way unified the various demographic groups of the population, which varied significantly along ethnic, religious, and cultural lines. The newfound government continued to evolve and suffer from inadequate infrastructure throughout the 20th century. Having only transitioned from military-style rule to civilian-led government in 1999, the country today still struggles with political instability, a lack of basic goods and services, and maintaining national rule of law.[59] 40% of the population today lives below the poverty line.[60]

Similarly, Nigeria continues to be one of Africa's most diverse countries with over 250 ethnic groups and more than 500 indigenous languages. There are 8 predominant ethnic groups (Hausa 30%, Yoruba 15.5%, Igbo/Ibo 15.2%, Fulani 6%, Tiv 2.4%, Kanuri/Beriberi 2.4%, Ibibio 1.8%, Ijaw/Izon 1.8%, other 24.7%, 2018 est.) and almost a 50-50 religious divide between Muslims and Christains (Muslim 53.5%, Christain 45.9%, other .6%, 2018 est.).[61] Unfortunately, these identifiers often become salient points of division over issues of resource sharing, making ethnicity and religion substantial conflict amplifiers.

One of the most violent problems of resource sharing has become the herder-farmer divide. Agriculture continues to be a dominant market in the Nigerian economy, responsible for over 20% of GDP[62] and supporting up to one-half of all Nigerian livelihoods.[63] Within this sector are both farming and livestock, which employ an overwhelming amount of Nigerian households. But these two activities tend to require the same resources — namely fertile land, water, and area. Normally, the two modes of agriculture would operate independently in coexistence. However, in recent years, the effects of climate change have proved especially devastating in northern Nigeria where desertification has made livestock cultivation near-impossible, pushing

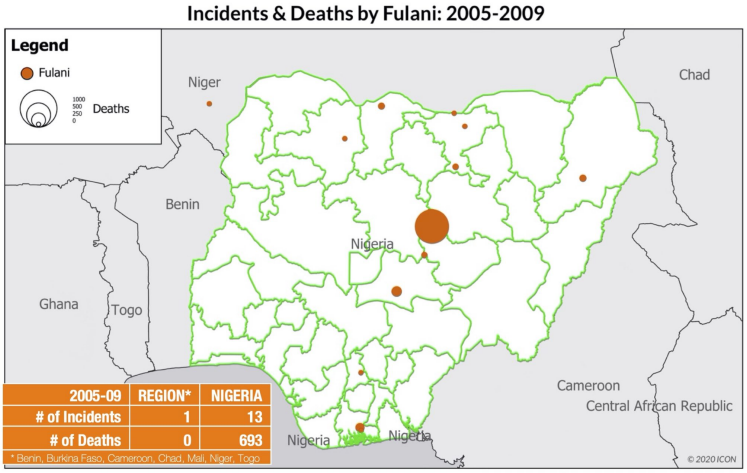
[57] "Nigeria: The World Factbook." *Central Intelligence Agency*, Central Intelligence Agency, <www.cia.gov/the-world-factbook/countries/nigeria/#people-and-society.> Accessed 30 Mar. 2021,
[58] Ochonu, Moses. "The Roots of Nigeria's Religious and Ethnic Conflict." *The World* , 10 Mar. 2014, www.pri.org/stories/2014-03-10/roots-nigerias-religious-and-ethnic-conflict.
[59] "Nigeria," CIA.
[60] "The World Bank In Nigeria." *World Bank*, <www.worldbank.org/en/country/nigeria/overview.>. Accessed 15 Mar 2021.
[61] "Nigeria," CIA.
[62] "Topic: Agriculture in Nigeria." *Statista*, <www.statista.com/topics/6729/agriculture-in-nigeria/.>. Accessed 15 Mar 2021.
[63] "Economy of Nigeria." *Encyclopædia Britannica*, Encyclopædia Britannica, Inc., <www.britannica.com/place/Nigeria/Economy.> Accessed 15 Mar 2021.

Fulani herdsmen southward in search of more suitable grazing land.[64] This has caused clashes between farmers and herdsmen competing over resources. Disputes over water pollution, cattle theft, and crop destruction have continued to increase, with disputes more often turning violent.

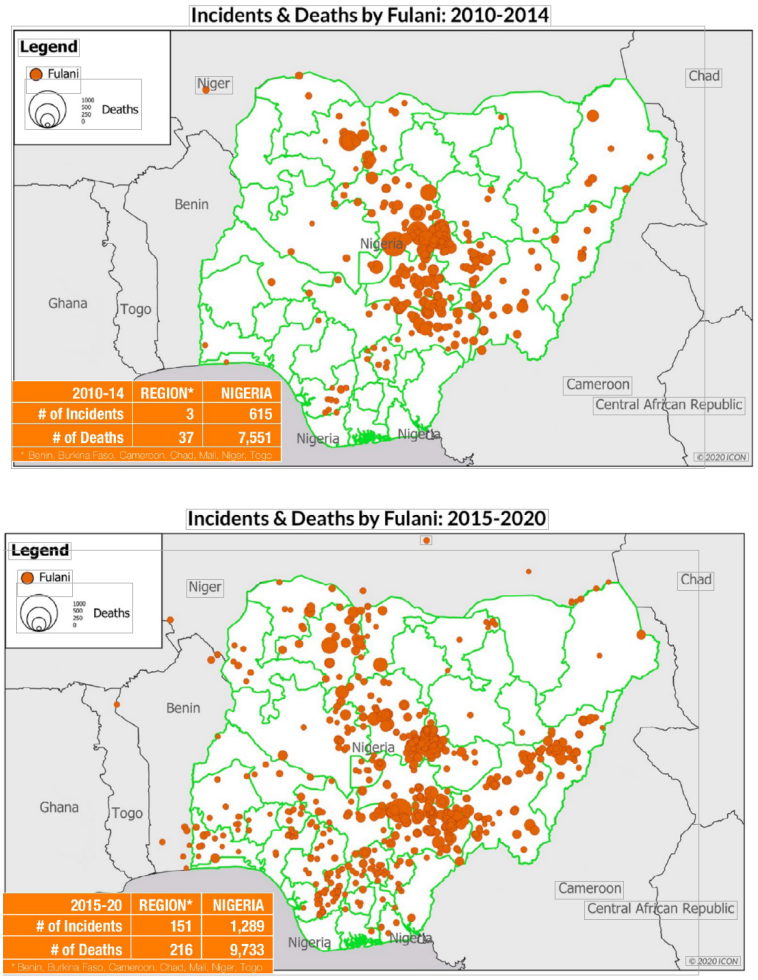
Since 2010, this conflict over shared resources has transformed into a conflict of another kind: ethno-religious targeted killings by and against groups of Fulani militants, driven by fears of scarce resources which are exacerbated by ethnic and religious divides. Militant tribal groups—including members of the Berom, Tarok, Eggon, and Jukun ethnic communities—would often form in efforts to confront Fulani herders, leading to targeted killings along ethnic lines on both sides.[65]

Overview of the Current State of the Conflict

Ethno-religious tribal violence in Nigeria has escalated significantly since late 2017 after the introduction of new anti-grazing laws that prohibited traditional herder practices of letting cattle roam freely.[66] These laws "outlaw[ed] the pastoralism practiced by many Fulani for generations" and sparked attacks by Bachama youth militias on Fulani herding communities, killing over 50 and causing retaliatory attacks by Fulanis the following month.[67] Since then, violence has only continued to escalate due to increasingly detrimental environmental conditions, the growth of local ethnic militia groups, insufficient government response, and additional open grazing bans. Officials have reported the violence incited by Fulani militants to be six timeless deadlier than Boko Haram, Africa's most infamous terrorist group, killing over 1,300 Nigerian civilians in the first half of 2018.[68] The humanitarian and economic tolls have been drastic, forcing over 300,000 children out of school in Benue State alone.[69]



[64] "Stopping Nigeria's Spiralling Farmer-Herder Violence." Crisis Group, 26 July 2018, www.crisisgroup.org/africa/west-africa/nigeria/262-stopping-nigerias-spiralling-farmer-herder-violence.
[65] Ibid.
[66] Ibid.
[67] Ibid.
[68] Ibid.
[69] Ibid.



Diagrams: Incidents & Deaths by Fulani, 2005-2020. Source: ICON PSJ

The current conflict, while perhaps having begun as competition over resources, has taken on an overtly ethnic dimension, in which Fulani groups are consistently either the perpetrator or the victim. Increasingly large and organized militias continue to form along ethnic lines, some of the most prominent being the Berom and Tarok of Plateau State; Eggon of Nasarawa State; and Jukun of Taraba State. The Fulani population, the majority of which is innocent of genocidal crimes, have become a collective target. Fulani youth group JAFUYAN commented on killings in the Numan area of Adamawa State, saying they were only "the latest in a coordinated agenda to wipe out our people systematically through ethnic cleansing."[70] These attacks often inevitably incite violence from the victims, in this case the Fulani, in acts of revenge. In the Benue state, His Royal Majesty James Ortese Ayatse, a Tiv tribal leader, described retaliatory attacks on his people as "well-planned … nothing short of ethnic cleansing and genocide against the Tiv nation."[71] There has been controversy about federal military support of attacks against certain

[70] "Stopping," 2018.
[71] Ibid.

groups: in March of 2018, retired lieutenant general Theophilus Danjuma accused the military of "ethnic cleansing" and called on citizens to defend themselves.[72] Recently, the situation has been termed a genocide by an increasing number of international organizations and human rights groups.[73]

Linking Misinformation to Recent Events

Nigeria is Facebook's largest and fastest growing African market with over 30 million users.[74] While this has allowed for tremendous operating growth, it has also created a breeding ground for rampant misinformation and disinformation that vilifies groups along ethnic lines. Misinformation is often spread unintentionally, as Facebook users routinely share posts relating to ethnic violence without realizing the full impact of the information they are sharing. However, the purposeful spreading of disinformation to inflame ethnic divides has become increasingly common. Ilia Djadi, formerly an employee of World Watch Monitor, described the ease with which media spreads across Nigeria as a reality that "opens doors to all manner of manipulation of data or facts."[75]

For example, just preceding the escalation of the conflict in 2017, national Nigerian media personality Linda Ikeji posted photos of young children killed during a clash between local Fulani and farming groups, describing the victims as children of farmers. Several comments attacked and vilified the Fulani herders as a result. However, the children themselves were not from herding communities: they were Fulani.[76] Uninformed postings such as these encourage sentiments of ethnic hatred while vilifying undeserving victims. However, more targeted and intentional postings often have more nefarious — and more lethal — real-time consequences.

On June 23, 2018, a picture of a baby hacked to death with machete marks on its face and a video of a man's head being cut open began circulating Nigeria's Facebook networks. There were accompanying images of bloody dead bodies in mass graves and children murdered in their homes. The user who posted the images attributed them to a killing spree happening in the Gashish district of Plateau State: Fulani Muslims killing Berom Christains. The post of the baby included a message to "wipe out the entire generation of the killers of this innocent child."[77]

---

[72] Ibid.

[73] Smith, Samuel. "Fulani Killings of Nigerian Christians Meets Standards for Genocide, Jubilee Campaign Says." *The Christian Post* via Genocide Watch, 22 Jul 2019. <https://www.genocidewatch.com/single-post/2019/08/01/fulani-killings-of-nigerian-christians-meets-standard-for-genocide-jubilee-campaign-says>.

[74] Oludimu, Titilola. "Facebook Now Has a Combined 70.9m Monthly Active Users in Nigeria, SA and Kenya." *Techpoint Africa*, 14 Oct. 2019, techpoint.africa/2019/10/14/facebook-nigeria-active-users/.

[75] *Nigeria: Unfolding Genocide? An Inquiry by the UK All-Party Parliamentary Group for International Freedom of Religion or Belief*. All Party Parliamentary Group for International Freedom of Religion or Belief, 9 Mar 2020. <https://appgfreedomofreligionorbelief.org/media/200615-Nigeria-Unfolding-Genocide-Report-of-the-APPG-for-FoRB.pdf>.

[76] Seip-Nuño, Ginger. *Vulnerable Narrative: Media Coverage of the Changing Pastoral Conflict in Nigeria*, Small Wars Journal, 2 Oct. 2018, <https://smallwarsjournal.com/jrnl/art/vulnerable-narrative-media-coverage-changing-pastoral-conflict-nigeria>.

[77] Adegoke, Yemisi. "Nigerian Police," 2018.

According to the post, Fulanis had just launched a targeted attack on the local Berom community. But these images had nothing to do with the recent attacks in Gashish—in fact, they had nothing to do with the conflict at all. The video of a man's skull being hacked open was recorded in Congo-Brazzaville almost a thousand miles away in 2012. The image of the baby appeared on Facebook months earlier.

The consequences of this disinformation were immediate and deadly. A Berom youth leader said in response to the images that were circulated, "As soon as we saw those images, we wanted to just strangle any Fulani man standing next to us… "Who would not, if they saw their brother being killed?"[78] The next day, a rally made up of members from the Berom ethnic community with machetes and knives blocked a local road and started interrogating drivers, searching out Fulani Muslims. 11 men were dragged from their vehicles and burned or violently hacked to death. In subsequent days, these bodies were found all over the city.[79]

The link here between Facebook's platforms and the attacks are undeniable. Plateau State public relations officer Tyopev Terna Matthias said the place of the incident "was not under attack [as alleged by the images]. But because of those images they [members of the Berom community] saw, the next day, roads were blocked. People died. Vehicles were burned. So many people died."[80] Officials pointed out that, aside from insufficient police response, another part of the problem was the fact that at the time, Facebook had employed a total of *four* fact-checkers for all of Nigeria — none of which spoke the local language of Hausa — for the over 24 million monthly users.

Non-locals have used Facebook to promote regional ethnic violence as well. Fulani Dr. Idris Ahmed is a prime example. Based in the UK, he repeatedly attempted to drive Fulanis to target Beroms in Nigeria through multiple postings on Facebook. One of his many postings:

[78] Ibid.
[79] Ibid.
[80] Ibid.

Source: BBC[81]

Dr. Ahmed's postings have fueled the momentum of multiple attacks on the Berom community. In response, Berom youth leader Joshua Pwajok called Ahmed "an element of destruction" who has "succeeded in building an army of hatred, an army among his people."[82] Professor Sam Godongs of Jos University in Nigeria condemned Dr. Ahmed's online activity, saying that "cumulatively, this is responsible for genocide."[83]

Summary of Trends

Misinformation on social media has amplified the potential for conflict both on behalf of and against the Fulani in Nigeria. Posts shared by influential public figures contributed to the fostering of widespread violent sentiment, and misinformation in the form of graphic videos (i.e. not direct calls to violence, but rather posts which masquerade as simply informational and appear to merit retaliation) provided the spark necessary for deadly outbreaks of violence.

[81] Ibid.
[82] Ibid.
[83] Ibid.

India

Background on Hindu Nationalism

With a population of nearly 1.4 billion people, India is one of the most culturally, linguistically, and racially diverse nations on the planet.[84] However, no ethnic cleavage has been quite as salient in India's politics and history as religion. The subcontinent is composed of several major religious groups, with around 80% of Indians identifying as Hindu, around 15% identifying as Muslim, and around 2% identifying as Christian and Sikh, each.[85] This fact of a significant majority Hindu population has created a dynamic in which minority religions face discrimination and violence in the name of "Hindu Nationalism." (This case study narrows its scope to focus on threats to Sikhs and Muslims, but violence against other religious minorities is present in India as well).

The ideology of Hindu Nationalism (sometimes referred to as "*Hindutva"*) centers around the belief that the state of India should be united and shaped by Hindu culture and religion. Hindu Nationalism is both a social and a political movement, and can be traced back to India's period of colonial rule by the British. Rallying around a Hindu identity was a form of backlash both against British occupation and against the Islamic dynasty that had ruled before that. In the 1920s, poorer/lower caste Hindus embraced Hindu nationalism more, as they felt oppressed by Hindu elites and were rallied around the idea of "returning India to its pure Hindu roots."[86] This is significant in that it equates being a true Indian with being Hindu, thus excluding members of other religions such as Sikhs and Muslims.

In 1925, while India was under British rule, a Hindu Nationalist organization called Rashtriya Swayamsevak Sangh (RSS) was founded to advocate for the creation of a Hindu nation. It differed from other nonviolent Hindu statist groups in that it emphasized Hindu scripture and military discipline.[87] In 1951, the Bharatiya Jana Sangh ("Indian People's Association") (BJS) was created as the political wing for the RSS, which considered itself a cultural/social organization. By 1967, BJS had substantial support in northern India. BJS joined forces with several other parties in 1977 to form a larger "Janata Party" which collapsed in 1979 due to factionalism. The Bharatiya Janata Party (BJP) was established in 1980 by a small group of dissidents, and has been running on a platform of Hindu nationalism since.[88]

---

[84] "India." *India - Place Explorer - Data Commons*, <https://datacommons.org/place/country/IND?utm_medium=explore&mprop=count&popt=Person&hl=en> Accessed 10 Mar 2021.

[85] "Religion Census 2011." Religion Data - Population of Hindu / Muslim / Sikh / Christian - Census 2011 India, <www.census2011.co.in/religion.php.>

[86] Shackle, Samira. ""What is Hindu nationalism?" *New Humanist,* The Rationalist Association, 27 May 2014. <https://newhumanist.org.uk/articles/4663/what-is-hindu-nationalism>.

[87] Frayer, Lauren, and Furkan Latif Khan. "The Powerful Group Shaping The Rise Of Hindu Nationalism In India." *NPR*, NPR, 3 May 2019, <www.npr.org/2019/05/03/706808616/the-powerful-group-shaping-the-rise-of-hindu-nationalism-in-india.>

[88] "Bharatiya Janata Party." *Encyclopædia Britannica*, Encyclopædia Britannica, Inc., <www.britannica.com/topic/Bharatiya-Janata-Party>.

The BJP's mainstream electoral success began in 1989, when the party called for the construction of a Hindu temple in an area occupied by a Mosque — demonstrating that from its very beginnings, the BJP's popularity was connected to capitalizing on Hindu nationalist sentiment. BJP party member Narendra Modi was the chief minister of the Gujarat state when a series of bloody riots, fueled by the destruction of a mosque and a train fire which led to the death of Hindu pilgrims, resulted in the murder of over 790 Muslims. Modi reportedly told state officials to take no action against the rioters, in effect allowing the massacre to continue.[89] In 2014, Modi was elected as Prime Minister and has since been advocating for policies that support a Hindu Nationalist agenda.[90]

Current State of the Ethno-Religious Conflict

India's move away from the secular state has had serious consequences for many of its citizens. This has been exemplified through anti-Muslim violence in the last several years and anti-Sikh sentiment during the 2020-2021 farmer's protest movement.

Discrimination Against Muslims

Between Modi's election to Prime Minister in 2014 and March 2019, there were 168 attacks against Muslims and other religious minorities in the name of Hindu Nationalism (many of them around protecting cows), killing 46 people. Communal violence has also risen 28% under Modi between the years 2014 and 2017. Some of this violence can be linked directly to BJP laws, such as an anti-conversion law that supposedly protects Hindus from the proselytizing of Muslims and has been cited by Bajrang Dal (a Hindu nationalist militant group) as justification for violence against these groups.[91] The BJP also took a step towards the official marginalization of Muslims in 2019 through the passage of a bill — the Citizenship Amendment Act — which requires a religious test for new immigrants, giving migrants of all major South Asian religions a path to citizenship except for those who practice Islam.[92] Such alienation has created the conditions for outbreaks of ethnic violence — which have resulted in death, torture, and property destruction, among other violations — via incendiary rumors and misinformation about Muslims on social media.

[89] "Timeline of the Riots in Modi's Gujarat." *The New York Times,* 19 Aug 2015. <https://www.nytimes.com/interactive/2014/04/06/world/asia/modi-gujarat-riots-timeline.html#/#time287_8192>.
[90] Frayer, Lauren. "Hindu Nationalism, The Growing Trend In India." *NPR*, NPR, 22 Apr. 2019, <www.npr.org/2019/04/22/715875298/hindu-nationalism-the-growing-trend-in-india.>
[91] Griswold, Eliza. "The Violent Toll of Hindu Nationalism in India. *The New Yorker,* 05 Mar 2019. <https://www.newyorker.com/news/on-religion/the-violent-toll-of-hindu-nationalism-in-india>.
[92] Gettleman, Jeffrey and Suhasini Raj. "India Steps Towards Making Naturalization Harder for Muslims." *The New York Times,* 9 Dec 2019. <https://www.nytimes.com/2019/12/09/world/asia/india-muslims-citizenship-narendra-modi.html>.

Sikh Farmers' Protests

A large proportion of India's workforce is engaged in agriculture: over 40% of the population works in an agricultural industry. However, farming contributes to only 17% of India's GDP, meaning the total economic output of the agricultural industry is low given how large of a sector it is. Farmers in India are, in general, struggling to make ends meet: over 20% of India's farmers live below the poverty line.[93] Such difficulties contribute to what has been termed "an epidemic of farmer suicides."[94]

In what they say is an effort to combat growing challenges for farmers, the Modi government passed 3 bills in September of 2020: The Farmers' Produce Trade and Commerce (Promotion and Facilitation) Act, the Farmers (Empowerment and Protection) Agreement on Price Assurance and Farm Services Act, and the Essential Commodities (Amendment) Act. Together, these 3 bills deregulate the agricultural market, remove stockpiling limits, and provide a legal framework for contract farming. Farmers can sell goods for any price to anyone.[95] According to Modi, this is beneficial to farmers in that they now have more freedom to sell to buyers directly without the use of a middleman and sell to other states and large grocery chains. However, farmers fear these bills will have significant negative consequences. The legislation will potentially allow big companies to drive prices down, such that farmers will struggle to meet the low price demand in years when the supply is high. Private buyers can sidestep mandis (agricultural markets) and coordinate lower prices. Farmers are also worried the government will dismantle the MPS system (minimum support price), which is currently in place to protect farmers in case the open market price gets too high.)

To protest these bills and force the government to repeal them, tens of thousands of farmers have flooded to New Delhi, the Indian capital, and have been camped out on highways in tents since November. It is important to note that many of these protesters are Sikhs hailing from the Punjab region, a heavily Sikh agricultural area in northern India. In the responses to these protestors, both from the Modi administration and from the online Hindu Nationalist community, dangerous anti-Sikh sentiment is apparent. The government has made continual attempts to silence and suppress protesters, sometimes at the cost of their lives. More than a hundred protestors have gone missing, and one woman activist was reported to be detained and sexually assaulted.[96] Many Sikhs fear that the clashes and the Hindu nationalist rhetoric in response to the protests will lead to a religious pogrom, similar to the likes of one which occurred in 1984 after two Sikh bodyguards assassinated then-Prime Minister Indira Gandhi in

[93] Menon, Shruti. "India Farmer Protests: How Rural Incomes Have Struggled to Keep Up." *BBC News*, BBC, 8 Feb. 2021, <www.bbc.com/news/world-asia-india-55413499.>

[94] Carleton, Tamma A. "Reply to Plewis, Murari et al., and Das: The suicide–temperature link in India and the evidence of an agricultural channel are robust." *Proceedings of the National Academy of Science of the United States America,* 9 Jan 2018. <https://www.pnas.org/content/115/2/E118>.

[95] Damodaran, Harish. "Explained: The Concerns of Farmers, and What Centre Can Negotiate to End Protests." *The Indian Express*, 15 Feb. 2021, <https://indianexpress.com/article/explained/farmers-big-concern-and-what-govt-could-negotiate-7073291/>

[96] Hundal, Sunny. "Why India's Farmers' Protests Have Sikhs Fearing Violent Attacks." OpenDemocracy, 4 February 2021. <www.opendemocracy.net/en/why-indias-farmers-protests-have-sikhs-fearing-violent-attacks/.>

retaliation for attacks on a Sikh holy site that killed thousands of civilians. The subsequent violence by Hindu mobs resulted in over 8,000 Sikhs dead in all of India — 3,000 in New Delhi alone — and is known by some, particularly within the Sikh community, as the 1984 Genocide.[97]

Linking Misinformation to Recent Events

Misinformation and Mob Violence Against Muslims

Since the introduction of social media, communal violence against Muslims has been sparked by rumors that spread through platforms like Facebook and WhatsApp. In 2013, the Muzzaffarnagar riots in the northern state of Uttar Pradesh, which left over 60 dead and thousands displaced, were triggered by a fake video circulating on social media, which was rumored to depict a Muslim mob brutally murdering a Hindu youth. This video was allegedly uploaded by a BJP legislator. In 2018, rumors online provoked several violent and sometimes fatal attacks in the last few years against Muslims purported to have killed or smuggled cows, which are sacred animals to Hindus.[98] Hindu nationalists take advantage of online platforms to sow fear and anger among the Hindu majority, alleging (falsely) that mosques have been built on destroyed temples, or that Muslims are trying to seize a majority over Hindus through "land jihad," "love jihad," and even "corona jihad."[99]

Misinformation and the Sikh Farmers' Protests

Misinformation, inflammatory messaging, and censorship on social media have also played a role in creating a threat for Sikhs during the farmers' protests. Media outlets have associated Sikh protesters with an extremist separatist movement which aims to establish a Sikh homeland called "Khalistan"[100] in Punjab, publishing stories which tell of a "Khalistan conspiracy" and claiming that the protests have been overrun with Sikh extremists.[101] The Indian central government has also parrotted this misinformation, using the presence of Sikhs at the protests as an excuse to call all of the protestors radical separatists and diminish their political credibility.[102]

---

[97] Jeet Singh, Simran. "It's Time India Accept Responsibility for Its 1984 Sikh Genocide." *TIME Magazine,* TIME USA, 31 Oct 2014. <https://time.com/3545867/india-1984-sikh-genocide-anniversary/>

[98] "India: Mob lynches Muslim man accused of smuggling." *DW,* 21 Jul 2018. <https://www.dw.com/en/india-mob-lynches-muslim-man-accused-of-cow-smuggling/a-44771278>.

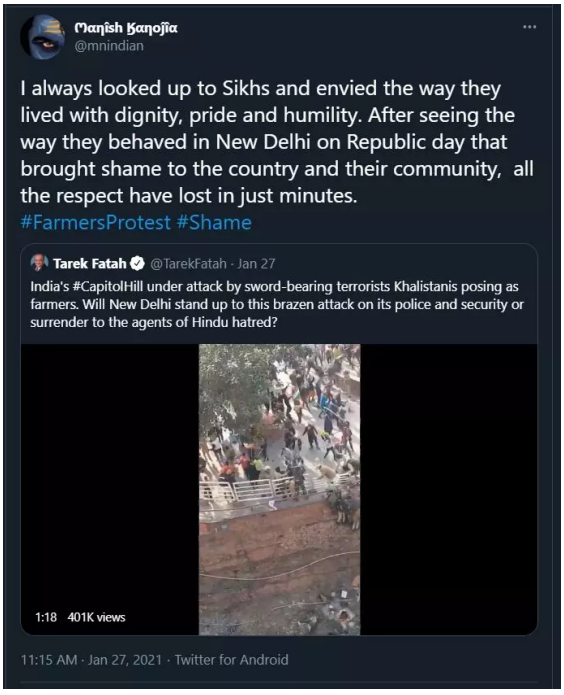
[99] Ali, Mohammad. "The Rise of a Hindu Vigilante in the Age of WhatsApp and Modi." *Wired,* 14 Apr 2020. <https://www.wired.com/story/indias-frightening-descent-social-media-terror/>.

[100] Britannica, The Editors of Encyclopaedia. "Khalistan". *Encyclopedia Britannica*, 18 Jun. 2009, <https://www.britannica.com/topic/Khalistan.>

[101] Bal, Hartosh Singh. "The New Khalistan Conspiracy." The Caravan, 27 Jan. 2021, <https://caravanmagazine.in/politics/republic-day-khalistan-conspiracy-government-playing-dangerous-game-punjab-disaster.>

[102] Khosla, Saksham and Aidan Milliff. "India's farm protests turned violent last week. But why are farmers protesting in the first place?" *The Washington Post,* 2021 Feb 5. <www.washingtonpost.com/politics/2021/02/05/indias-farm-protests-turned-violent-last-week-why-are-farmers-protesting-first-place/.>

Pictures and videos of pro-Khalistan rallies began to circulate in late 2020 as evidence of the separatist nature of the farmers' protests, but later fact-checking revealed that these videos were actually taken during separate rallies in different years.[103]



[103] Alphonso, Anmol. "Farmers' Protests: How Old Visuals Were Used to Peddle A Pro-Khalistan Narrative." *BOOM Fact Check,* BOOM, 11 Dec 2020. <https://www.boomlive.in/fake-news/farmers-protests-how-old-visuals-were-used-to-peddle-a-pro-khalistan-narrative-11090>.

Images: Example of misinformation linking the Farmers' Protests to the Khalistan separatist movement. Source: BOOM

Following the historical relationship between the Sikh separatist movement and the assassination of Prime Minister Indira Gandhi, posts containing the words "genocide" and "1984" and hashtags including #missingIndira and #shoot began to trend on Facebook and Twitter.[104] Clearly, these tweets are a direct incitement of violence against Sikh people.

Source: BOOM

Twitter reportedly took actions to censor, flag, and delete many of these posts, and suspended more than 500 accounts.[105] However, censorship was going in both directions. On January 26th, farmers who has been protesting clashed with police, and discourse on Twitter surrounding the event became heated with some accusing Modi of carrying out a genocide against the farmers. Modi drew upon an Indian law that allows governments to issue a blocking order on tech platforms to prevent inciting violence and had Twitter agree to block around 250 accounts (including The Caravan, a popular Indian publication). Many of these accounts were using the hashtag #modiplanningfarmergenocide. Many Indian people protested this, saying that this was a violation of free speech, but the Modi government declared non-compliance with an order when Twitter reinstated the accounts. Twitter temporarily muted some accounts and hashtags, although all accounts were eventually restored.[106] This is an example of the Modi government using social media as a political tool to silence criticism in an environment already scarred by divisive misinformation.

---

[104] Alphonso, Anmol. "'Repeat 1984' 'Missing Indira Gandhi': Chilling Anti-Sikh Tweets Rise." *BOOM Fact Check,* BOOM, 29 Jan 2021. <https://www.boomlive.in/politics/farmers-protests-anti-sikh-tweets-twitter-republic-day-tractor-rally-hate-speech-violence-delhi-khalistan-rhetoric-1984-riots-11761>

[105] Sircar, Sushovan. "Suspended 500 Accounts: Twitter on Violent Calls to 'Repeat 1984'." *TheQuint*, 28 Jan. 2021, <www.thequint.com/cyber/twitter-500-accounts-suspended-violent-calls-repeat-of-1984-red-fort-sikh-farmers.>

[106] Frayer, Lauren and Shannon Bond. "Twitter in Standoff With India's Government Over Free Speech and Local Law." *NPR Technology,* National Public Radio, 18 Feb 2021. <https://www.npr.org/2021/02/17/968641246/twitter-in-standoff-with-indias-government-over-free-speech-and-local-law>

Summary of Trends

Misinformation on social media has amplified the potential for conflict on behalf Hindu nationalists and against ethno-religious minorities in India. In the case of Muslim Indians, misinformation in the form of graphic videos which characterize Muslims as a threat combined with the history of state-sanctioned Hindu nationalist violence and discrimination against Muslims provided the spark necessary for deadly outbreaks of mob violence. In this way, misinformation and outcomes of ethnic violence may be seen as linked. In the case of Sikh Indians, misinformation in the form of charges of Khalistani terrorism — which still lingers in the national imagination as a justifiable cause for mass violence — has increased direct hate speech against Sikhs. In this way, misinformation can lead to online speech which counts as incitement to violence (a violation of ICCPR Article 20(b)).

Framework

The following six-pronged framework of inquiry may be employed to define misinformation which is incitement to ethnic violence. This framework is adapted from work[107] by human rights and dangerous speech expert Susan Benesch [link] which aims to guide governing bodies in determining whether an individual has committed incitement to genocide — a charge which is similar to misinformation in that it is nebulous and inextricably linked to the question of free speech. Our framework is not intended for prosecutorial usage in the same sense (i.e. criminal charges against a specific actor), but rather to judge the content of a particular post. It is informed by trends observed and analyzed in the three case studies.

Framework on Defining Misinformation as Incitement to Violence

1. **Was the speech understood by the audience as a call to ethnic violence?**
    a. In each case study, major instances of ethnic violence are linked to misinformation that includes a graphic video or image of members in one ethnic group purportedly causing bodily harm to members in another ethnic group ("Imagery of Harm"), even when the attribution of said acts to a particular ethnic group turns out to be false (either in terms of context or in the actors identified).
    b. Even where such Imagery of Harm posts do not include a specific call to violence, the incendiary nature of this graphic imagery combined with the mis-attribution to an already-feared or disliked ethnic group often leads to the interpretation that violence is a necessary response.
    c. Special attention should thus be paid to images or posts which appear to depict members of one ethnic group harming members of another ethnic group, as misinformation in this form (misattribution of graphic activity) has a demonstrated capacity to incite violence.
2. **Did the post reach a large audience and did the audience have the capacity to commit violence?**
    a. Using the concept of a "large audience" accounts for situations in which the person sharing or creating the post has a large following (i.e. public figures like beauty queen Shwe Eain Si in the Rohingya/Myanmar case study, Dr. Idris Ahmed in the Fulani/Nigeria case study) *and* situations where the post itself has gone viral through interpersonal sharing, as people tend to trust posts shared by people they trust (i.e. rumors about Muslims circulated via WhatsApp messaging in the India case study).

[107] We thank Professor Benesch for providing the inspiration for this framework. We also note that this particular work was published in 2008, and that our understanding of dangerous speech has grown since then. See dangerousspeech.org for more.

   b. The capacity to commit violence in our Myanmar and India case studies is indicated in that victims often belong to an ethno-religious minority while perpetrators belong to an ethno-religious majority with state power. In the Nigeria case study, the capacity to commit violence both on behalf of and against the Fulani is demonstrated in the militant nature of the tribal groups involved, as they engaged in armed conflict with one another prior to the introduction of social media.

3. **Had the victims-to-be already suffered an outbreak of recent violence? Did the perpetrators-to-be recently commit acts of violence against minority ethnic groups?**
   a. Each of our three case studies noted a salient ethnic division that had resulted in acts of violence / violent clashes immediately prior to the introduction of social media. In Myanmar and Nigeria in particular, the situation demonstrably worsened after the introduction of social media (see timeline in the Myanmar case study and diagrams of increasing ethnic violence in the Nigeria case study).
4. **What did the platform for contrasting viewpoints look like? Were contrasting posts receiving as much engagement as their counterparts?**
   a. Another element to consider is that in Myanmar and India, state governments interfered with the viewpoints available to citizens through:
      i. members of the military posing as civilians to discredit foreign news services like BBC and drive up engagement on anti-Rohingya posts (Myanmar)
      ii. government officials requesting that criticisms of the farm bills and warnings of a farmer genocide be blocked (India)
   b. Special attention should thus be paid to instances where the state government also has expressed or demonstrated a stake in the ethnic divide, as this may increase censorship (direct or indirect via "drowning out" opposing viewpoints according to engagement algorithms).
5. **Did it use language or imagery, explicit or coded, to justify and promote violence? Did the post describe the victims-to-be as subhuman, or accuse them of perpetrating violence? Did it also make explicit or coded reference to their ethnicity? Had the audience been conditioned by the use of these techniques in other, previous speech?**
   a. As discussed in Point #1, all three case studies included instances where ethnic violence was directly linked to an Imagery of Harm post which included references to the victims' ethnicity.
   b. In addition to these major Imagery of Harm posts, all three case studies included posts where the victims-to-be were labeled "terrorists" and accused of perpetrating or plotting violence even when a specific video was not included.
   c. In the case of Myanmar and India, this mirrored government rhetoric.
   d. Such posts are thus also central to the psychological conditioning of misinformation which results in ethnic violence.

6. **Had the audience received similar messages before the post? Did the post appear often in tandem with other flagged posts, creating an echo chamber effect?**
    a. It is important to note that in every case study, one Imagery of Harm post was not single-handedly responsible for an outbreak of ethnic violence. Rather, the impact of such posts must be considered in tandem with an onslaught of other posts which are less graphic but contribute to making ethnic divisions salient. The nature of posts which would qualify under this definition is explored in Policy Recommendation III.

Policies

We agree with Facebook's assessment that online speech should not be proactively restricted or prohibited. Instead, we propose the following human rights-based approach and policy framework to promote more targeted review, removal, and informative practices regarding misinformation and ethnic violence:

I. Determining Which Posts to Monitor
II. Guidance on Third-Party Involvement
III. Proposed Criteria for Action (Flagging and Removal)
IV. Ethnic Violence Information Center

I. Determining Which Posts to Monitor

1. In the current Covid-19 pandemic, Facebook and its affiliates use AI to monitor coronavirus-related posts and attach a link which directs users in the United States to an informational site by the Centers for Disease Control (CDC)
2. We propose that Facebook use the same AI capacity to monitor posts regarding ethnic conflict.
3. This may be conducted in the following form:
    a. If the United Nations (e.g. Office of the High Commissioner for Human Rights), Genocide Watch, Human Rights Watch, Amnesty International, or other trusted human rights organization releases a statement regarding human rights violations in a particular ethnic conflict (e.g. the 2017 OHCHR statement on Myanmar as a "textbook example of ethnic cleansing"), Facebook AI must monitor and review posts related to said conflict.
        i. To determine which keywords and symbols the AI should use to identify relevant posts, Facebook should partner with advocacy organizations on the ground in said region (often identified in UN / human rights organization releases).
    b. If AI detects a serious uptick in mentions of a minority ethnic group in a particular country/region, Facebook should monitor and review posts related to said ethnic group.
        i. A starting point to classify a "minority ethnic group" in a particular country/region may be the World Directory of Minorities and Indigenous Peoples [link] or a similar resource.
    c. If AI detects the presence of a graphic video in tandem with the explicit or coded mention of a minority ethnic group, Facebook should place this at a high priority for review. "Imagery of Harm" misinformation posts tend to be particularly crucial in the incitement of ethnic violence.

II. Guidance on Third Party Involvement

1. To avoid being "the arbiter of truth", Facebook outsources fact-checking to vetted third partners. A crucial component to this fact-checking process is being able to fully understand content posted.
2. Facebook's policy statements on fact-checking say that fact-checkers are required to "review content, check its facts, and rate its accuracy" which can involve "calling sources, consulting public data, [or] authenticating videos and images."[108]
3. Content flagged and brought into question regarding incitement of violence through misinformation or disinformation will often be in the native language of its country of origin, or contain references to ethnic symbols or terms which are specific to its country of origin. As such, in order for fact-checkers to accurately perform their tasks as stated in Facebook's policy guidelines and communicate with outside or local officials and sources to better judge the content in question, it is imperative that they have a working knowledge of the language.
4. Thus we recommend that in team assignments of fact-checkers for various nations, Facebook enforce a quota system in which 25% of the fact-checking staff for a given country be fluent in the local language. In a multilingual state, this may be further subdivided to reflect the breakdown of major languages.

[108] "Fact-Checking on Facebook." *Facebook for Business,* Facebook. <https://www.facebook.com/business/help/2593586717571940?id=673052479947730>. Accessed 30 Mar 2021.

III. Proposed Criteria for Action (Flagging and Removal)

In this section, the group that the mis/disinformation may lead to commit violence will be called the PERPETRATORS, and the group that is on the receiving end of violence will be called the VICTIMS. Below, we propose four actions and their corresponding criteria.

Posts

Removal of a Post

Criteria for a post to warrant removal:

1. Contents of the post:
    a. The victims have been identified through explicit or coded terminology.
    b. 10%[109] of the comments under the post show that it has been interpreted by the perpetrators as a call to violence against the victims.
    c. The post itself contains incorrect information about the victims, including allegations that the victims have committed violence against the perpetrators.
2. Situation in the state:
    a. The victims have already been subjected to human rights violations by the perpetrators/government within the past three years.
    b. The government/or news outlets are censoring advocacy for the victims, either directly or indirectly via "drowning out" opposing viewpoints according to engagement algorithms.
    c. Influential figures (ex. religious authorities, government/military officials, popular culture influencers) have referred to the victims as subhuman, accused them of plotting violence/terrorism, or called for violence against them within the last two months.
3. Situation on Facebook:
    a. Advocacy groups for the victims are receiving significantly fewer impressions on Facebook (less than ⅓ the amount of impressions the perpetrators are receiving).
    b. There have been a significant quantity of posts on Facebook (as determined by AI metrics) within the last two months that describe the victims to be subhuman or terrorists.

ALL of these criteria must be fulfilled for a post to be removed. This is because all criteria are based on the above framework of inquiry, which in turn is based on Susan Benesch's 6-pronged definition of incitement of genocide, which in turn requires all 6 prongs to fulfilled in order to count as genocide. Upon the removal of a post, the user should receive notice that their post has

[109] Recognizing that Facebook already employs metrics for content moderation, and not knowing what those metrics are ourselves, we leave the exact number up to the discretion of your team.

been removed along with reasoning for why and a request that they do not continue to post such content, with a reminder of Facebook's community standards. This may include a description of whatever conflict they're contributing to along with resources on the situation (see Recommendation IV: Ethnic Violence Information Center).

Removal in the case of verified misinformation and unverifiable rumors which may result in violence is aligned with Facebook's new Corporate Human Rights Policy, as reported in the March 16 press release by Director of Human Rights Miranda Sissons.[110]

Flagging a Post

Criteria for a post to be flagged as "potentially contributing to ethnic violence":

1. Four or more but less than eight of the criteria in the removal section. At a minimum, this means one of the two criteria under "Contents of the Post," two of the four criteria under "Situation in the State," and one of the two criteria under "Situation on Facebook."

Upon the flagging of a post, the user should receive notice that their post has been flagged along with reasoning for why. This may include a description of whatever conflict they're contributing to along with resources on the situation.

Users

In cases of sustained misinformation, banning a user may be necessary to mitigate the likelihood of incitement to ethnic violence.

Banning a User Temporarily

Criteria for a user to be removed from the platform for three weeks:

1. The user has posted:
    a. five flagged posts or
    b. two removed posts or
    c. three flagged posts and one removed post.

Upon being banned temporarily, the user should receive a clear explanation for why, followed by an agreement they can sign for reinstatement in three weeks' time provided they agree not to continue posting such content (by signing an updated version of the terms of service) or risk permanent deactivation.

Banning a User Permanently

Criteria for a user to be removed from the platform permanently:

[110] Sissons, "Human Rights," 2021.

1. The user has posted:
    a. three more flagged posts or
    b. one removed post since their account's reinstatement.

IV. Ethnic Violence Information Center

1. Mirroring Facebook's current policy to direct Covid-19 related posts in the United States to the CDC, we propose that posts related to ethnic violence (identified using the guidance in Recommendation 1: Determining Which Posts to Monitor) also be directed to resources about said conflict.
2. These resources may be compiled by the same organizations which inform the monitoring of said content (e.g. United Nations, Human Rights Watch, etc.) and local advocacy groups.
3. In line with Recommendation II: Guidance on Third Party Involvement, such resources should also appear in the native language(s) of the country or region in question.
4. Employing the resources of intergovernmental organizations (IGOs) and nongovernmental organizations (NGOs) who specialize in human rights will help Facebook employ a rights-based approach to informing users about misinformation.

Sources

List of Experts Consulted

- Professor Jacqueline Bhabha, Professor of the Practice of Health and Human Rights at Harvard Law School
- Professor Susan Benesch, Executive Director of the Dangerous Speech Project and Fellow at the Harvard University Berkman Klein Center for Internet & Society
- Text of the International Covenant on Civil and Political Rights (ICCPR), the International Covenant on Economic, Social, and Cultural Rights (ICESCR), and the Convention on the Prevention and Punishment of the Crime of Genocide
- Press releases and statements from the United Nations, including the Office of the High Commissioner for Human Rights

Bibliography

"Addressing the Human Rights Impact of Misinformation Without Further Harming Human Rights." HRBDT, The Human Rights, Big Data and Technology Project. <https://www.hrbdt.ac.uk/addressing-the-human-rights-impacts- of-misinformation/> Accessed 23 Feb 2021.

Adegoke, Yemisi. "Nigerian Police Say 'Fake News' on Facebook Is Killing People." BBC News, BBC, 13 Nov 2018. <www.bbc.co.uk/news/resources/idt-sh/nigeria_fake_news.>

Al Jazeera Staff. "Who are the Rohingya?" Al Jazeera, Al Jazeera Media Network, 18 Apr 2018. <https://www.aljazeera.com/features/2018/4/18/who-are-the-rohingya>.

Ali, Mohammad. "The Rise of a Hindu Vigilante in the Age of WhatsApp and Modi." Wired, 14 Apr 2020. <https://www.wired.com/story/indias-frightening-descent-social-media-terror/>.

Alphonso, Anmol. "'Repeat 1984' 'Missing Indira Gandhi': Chilling Anti-Sikh Tweets Rise." BOOM Fact Check, BOOM, 29 Jan 2021. <https://www.boomlive.in/politics/farmers-protests-anti-sikh-tweets-twitter-republic-day-tractor-rally-hate-speech-violence-delhi-khalistan-rhetoric-1984-riots-11761>

Alphonso, Anmol. "Farmers' Protests: How Old Visuals Were Used to Peddle A Pro-Khalistan Narrative." BOOM Fact Check, BOOM, 11 Dec 2020. <https://www.boomlive.in/fake-news/farmers-protests-how-old-visuals-were-used-to-peddle-a-pro-khalistan-narrative-11090>.

Arakan Rohingya National Organization <https://www.rohingya.org/>.

Bal, Hartosh Singh. "The New Khalistan Conspiracy." The Caravan, 27 Jan. 2021,

<https://caravanmagazine.in/politics/republic-day-khalistan-conspiracy-government-playing-dangerous-game-punjab-disaster.>

BBC News Staff. "Myanmar: What sparked latest violence in Rakhine?" BBC News Asia, BBC, 19 Sept 2017. <https://www.bbc.com/news/world-asia-41082689>

Benesch, Susan. "But Facebook's Not a Country: How to Interpret Human Rights Law for Social Media Companies." Yale Journal on Regulation, 14 September 2020. <https://www.yalejreg.com/bulletin/but-facebooks-not-a-country-how-to-interpret-human-rights-law-for-social-media-companies/>.

Benesch, Susan. "Vile Crime or Inalienable Right: Defining Incitement to Genocide." Virginia Journal of International Law, 48(3), 2008.

"Bharatiya Janata Party." Encyclopædia Britannica, Encyclopædia Britannica, Inc., <www.britannica.com/topic/Bharatiya-Janata-Party>.

Britannica, The Editors of Encyclopaedia. "Khalistan". Encyclopedia Britannica, 18 Jun. 2009, <https://www.britannica.com/topic/Khalistan.>

Burmese Rohingya Organization UK. "Myanmar's 1982 Citizenship Law and Rohingya." Burma Campaign, Dec 2014. <https://www.burmacampaign.org.uk/media/Myanmar%E2%80%99s-1982-Citizenship- Law-and-Rohingya.pdf>.

Campbell, Alex. "How Data Privacy Laws Can Fight Fake News." *Just Security,* Reiss Center on Law and Security at New York University School of Law, 15 Aug 2019. <https://www.justsecurity.org/65795/how-data-privacy-laws-can-fight-fake-news/>.

Carleton, Tamma A. "Reply to Plewis, Murari et al., and Das: The suicide–temperature link in India and the evidence of an agricultural channel are robust." Proceedings of the National Academy of Science of the United States America, 9 Jan 2018. <https://www.pnas.org/content/115/2/E118>.

Coldewey, Devin. "Who Regulates Social Media?" TechCrunch, Verizon Media, 19 Oct 2020. <https://social.techcrunch.com/2020/10/19/who-regulates-social-media/>.

Damodaran, Harish. "Explained: The Concerns of Farmers, and What Centre Can Negotiate to End Protests." The Indian Express, 15 Feb. 2021, <https://indianexpress.com/article/explained/farmers-big-concern-and-what- govt-could-negotiate-7073291/>

Doctors Without Borders Staff. "Timeline: A visual history of the Rohingya refugee crisis." Doctors Without Borders News & Stories, Medecins Sans Frontieres / Doctors Without Borders, 21 Aug 2020. <https://www.doctorswithoutborders.org/what-we-do/news-stories/news/timeline-visual-history- rohingya-refugee-crisis>.

"Economy of Nigeria." Encyclopædia Britannica, Encyclopædia Britannica, Inc., <www.britannica.com/place/Nigeria/Economy.> Accessed 15 Mar 2021.

"Ethnic Cleansing." UN Office on Genocide Prevention and the Responsibility to Protect, United

Nations. <https://www.un.org/en/genocideprevention/ethnic-cleansing.shtml>. Accessed 07 Mar 2021.

"Fact-Checking on Facebook." Facebook for Business, Facebook. <https://www.facebook.com/business/help/ 2593586717571940?id=673052479947730>. Accessed 30 Mar 2021.

Frankel, Rafael (Facebook Director of Policy, APAC Emerging Countries). "An Update on the Situation in Myanmar." *Facebook Newsroom,* Facebook, 11 Feb 2021. <https://about.fb.com/news/2021/02/an-update-on-myanmar/>.

Frayer, Lauren and Shannon Bond. "Twitter in Standoff With India's Government Over Free Speech and Local Law." NPR Technology, National Public Radio, 18 Feb 2021. <https://www.npr.org/2021/02/17/968641246/twitter-in-standoff-with-indias-government-over-free-speech-and-local-law>

Frayer, Lauren, and Furkan Latif Khan. "The Powerful Group Shaping The Rise Of Hindu Nationalism In India." NPR, NPR, 3 May 2019, <www.npr.org/2019/05/03/706808616/the-powerful-group-shaping-the-rise-of-hindu -nationalism-in-india.>

Frayer, Lauren. "Hindu Nationalism, The Growing Trend In India." NPR, NPR, 22 Apr. 2019, <www.npr.org/2019/04/22/715875298/hindu-nationalism-the-growing-trend-in-india.>

Gettleman, Jeffrey and Suhasini Raj. "India Steps Towards Making Naturalization Harder for Muslims." The New York Times, 9 Dec 2019. <https://www.nytimes.com/2019/12/09/world/asia/india-muslims-citizenship- narendra-modi.html>.

Global Campaign for Free Expression. (n.d.). Towards an interpretation of article 20 of the ICCPR: Thresholds for the prohibition of incitement to hatred Work in Progress. The regional expert meeting on article 20, Organized by the Office of the High Commissioner for Human Rights,. <https://www.ohchr.org/Documents/Issues/Expression/ICCPR/Vienna/CRP7Callamard.pdf>

Griswold, Eliza. "The Violent Toll of Hindu Nationalism in India. The New Yorker, 05 Mar 2019. <https://www.newyorker.com/news/on-religion/the-violent-toll-of-hindu-nationalism-in-india>.

Habib, Mohshin. and Jubb, Christine. and Ahmad, Salahuddin. and Rahman, Masudur. and Pallard, Henri. and Ontario International Development Agency, issuing body. Forced migration of Rohingya : the untold experience / Mohshin Habib, Christine Jubb, Salahuddin Ahmad, Masudur Rahman, Henri Pallard; photographs, Salahuddin Ahmad Ontario International Development Agency, Canada Ottawa, Ontario 2018.

Human Rights Watch. "'All You Can Do is Pray: Crimes Against Humanity and Ethnic Cleansing of Rohingya Muslims in Burma's Arakan State." Human Rights Watch, 22 Apr 2013. <https://www.hrw.org/report/2013/04/22/all-you-can-do-pray/crimes-against-humanity-and-ethnic-cleansing-rohingya-muslims>.

Human Rights Watch. "Questions and Answers on Gambia's Genocide Case Against Myanmar before the International Court of Justice." Human Rights Watch, 5 Dec 2019. <https://www.hrw.org/news/2019/12/05/questions-and-answers-gambias-genocide-case-against-myanmar-international-court#>.

Hundal, Sunny. "Why India's Farmers' Protests Have Sikhs Fearing Violent Attacks." OpenDemocracy, 4 February 2021. <www.opendemocracy.net/en/why-indias-farmers-protests-have-sikhs-fearing-violent-attacks/.>

"India: Mob lynches Muslim man accused of smuggling." DW, 21 Jul 2018. <https://www.dw.com/en/india-mob-lynches-muslim-man-accused-of-cow-smuggling/a-44771278>.

"India." India - Place Explorer - Data Commons, <https://datacommons.org/place/country/IND?utm_medium=explore&mprop=count&popt=Person&hl=en> Accessed 10 Mar 2021.

International Crisis Group. "Report No. 290: Buddhism and State Power in Myanmar." International Crisis Group, 5 Sept 2017. <https://www.crisisgroup.org/asia/south-east-asia/myanmar/290-buddhism-and-state-power- myanmar>

"International Law." Burma Link, 27 Oct 2014. <https://www.burmalink.org/background/burma/international-crimes-and-impunity/international-law/#:~:text=While%20not%20a%20treaty%2C%20UDHR,Declaration%20on%20December%2010%2C%201948>.

Jalali, Rita, and Seymour Martin Lipset. 1992. "Racial and Ethnic Conflicts: A Global Perspective." Political Science Quarterly 107(4): 595.

Jeet Singh, Simran. "It's Time India Accept Responsibility for Its 1984 Sikh Genocide." TIME Magazine, TIME USA, 31 Oct 2014. <https://time.com/3545867/india-1984-sikh-genocide-anniversary/>

Kaye, David (UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression). "Mandate of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression." Office of the Special Rapporteur, United Nations, 3 Apr 2018. <https://www.lawyersforliberty.org/wp-content/uploads/2018/04/OL_MYS_03.04.18-1.pdf>

Khosla, Saksham and Aidan Milliff. "India's farm protests turned violent last week. But why are farmers protesting in the first place?" The Washington Post, 2021 Feb 5. <www.washingtonpost.com/politics/2021/02/05/indias-farm-protests-turned-violent-last-week-why-are-farmers-protesting-first-place/.>

Koch, Richie. "The GDPR meets its first challenge: Facebook." GDPR EU, Proton Technologies AG. <https://gdpr.eu/the-gdpr-meets-its-first-challenge-facebook/>. Accessed 07 Mar 2021.

"Malaysia: Anti-Fake News Act Comes Into Force." Global Legal Monitor, Library of Congress, 19 Apr 2018. <https://www.loc.gov/law/foreign-news/article/malaysia-anti-fake-news-act-comes-into-force/>

Menon, Shruti. "India Farmer Protests: How Rural Incomes Have Struggled to Keep Up." BBC News, BBC, 8 Feb. 2021, <www.bbc.com/news/world-asia-india-55413499.>

Mon, Ye. "Residents Flee, police move in after mob violence targets Muslims in Bago village." Myanmar Times, The Myanmar Times, 27 Jun 2016. <http://www.mmtimes.com/index.php/national-news/21055-residents-flee-police-move-in-after-mob-violence-targets-muslims-in-bago-village.html>.

Mozur, Paul. "A Genocide Incited on Facebook, with Posts from Myanmar's Military." The New York Times, 15 Oct 2018. <https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html>.

"Nigeria: The World Factbook." Central Intelligence Agency, Central Intelligence Agency, <www.cia.gov/the-world-factbook/countries/nigeria/#people-and-society.> Accessed 30 Mar. 2021,

Nigeria: Unfolding Genocide? An Inquiry by the UK All-Party Parliamentary Group for International Freedom of Religion or Belief. All Party Parliamentary Group for International Freedom of Religion or Belief, 9 Mar 2020. <https://appgfreedomofreligionorbelief.org/media/200615-Nigeria-Unfolding-Genocide-Report-of-the-APPG-for-FoRB.pdf>.

Ochonu, Moses. "The Roots of Nigeria's Religious and Ethnic Conflict." The World , 10 Mar. 2014, <www.pri.org/stories/2014-03-10/roots-nigerias-religious-and-ethnic-conflict.>

Persaud, Santhosh. "Protecting Refugees and Asylum Seekers under the International Covenant on Civil and Political Rights."New Issues in Refugee Research, UNHCR, Research Paper No. 132, Nov 2006. <https://www.unhcr.org/research/working/4552f0d82/protecting-refugees-asylum-seekers-under-international-covenant-civil-political.html>.

Persio, Sofia Lotto. "Why Myanmar Hates the Rohingya." Newsweek World, Newsweek Digital, 15 Sept 2017. <https://www.newsweek.com/why-myanmar-hates-rohingya-665883>.

Rajgopalan, Megha. "Internet Trolls Are using Facebook To Target Myanmar's Muslims." BuzzFeed News, BuzzFeed, 18 Mar 2017. <https://www.buzzfeednews.com/article/meghara/how-fake-news-and-online-hate- are-making-life-hell-for#.wlGyPB4gk>.

"Religion Census 2011." Religion Data - Population of Hindu / Muslim / Sikh / Christian - Census 2011 India, <www.census2011.co.in/religion.php.>

Reuters Staff. "Timeline: Three years on, a look at the Rohingya crisis." Reuters, 20 Aug 2020. <https://www.reuters.com/article/us-myanmar-rohingya-timeline/timeline-three-years-on-a-look-at-the-rohingya-crisis-idUSKBN25H03Y>

Seip-Nuño, Ginger. Vulnerable Narrative: Media Coverage of the Changing Pastoral Conflict in Nigeria, Small Wars Journal, 2 Oct. 2018, <https://smallwarsjournal.com/jrnl/art/vulnerable-narrative-media-coverage- changing-pastoral-conflict-nigeria>.

Shackle, Samira. ""What is Hindu nationalism?" New Humanist, The Rationalist Association, 27 May 2014. <https://newhumanist.org.uk/articles/4663/what-is-hindu-nationalism>.

Sircar, Sushovan. "Suspended 500 Accounts: Twitter on Violent Calls to 'Repeat 1984'." TheQuint, 28 Jan. 2021, <www.thequint.com/cyber/twitter-500-accounts-suspended-violent-calls-repeat-of-1984-red-fort-sikh-farmers.>

Sissons, Miranda. "Our Commitment to Human Rights." Facebook Newsroom, Facebook, 16 Mar 2021. <https://about.fb.com/news/2021/03/our-commitment-to-human-rights/>.

Smith, Samuel. "Fulani Killings of Nigerian Christians Meets Standards for Genocide, Jubilee Campaign Says." The Christian Post via Genocide Watch, 22 Jul 2019. <https://www.genocidewatch.com/single-post/2019/08/01/fulani-killings-of-nigerian-christians-meets-standard-for-genocide-jubilee-campaign-says>.

"Stopping Nigeria's Spiralling Farmer-Herder Violence." Crisis Group, 26 July 2018, <www.crisisgroup.org/africa/west-africa/nigeria/262-stopping-nigerias-spiralling-farmer-herder-violence.>

"The World Bank In Nigeria." World Bank, <www.worldbank.org/en/country/nigeria/overview.>. Accessed 15 Mar 2021.

"Timeline of the Riots in Modi's Gujarat." The New York Times, 19 Aug 2015. <https://www.nytimes.com/interactive/2014/04/06/world/asia/modi-gujarat-riots-timeline.html#/#time287_8192>.

"Topic: Agriculture in Nigeria." Statista, <www.statista.com/topics/6729/agriculture-in-nigeria/.>. Accessed 15 Mar 2021.

United Nations General Assembly. Convention on the Prevention and Punishment of the Crime of Genocide, 9 December 1948. <https://www.un.org/en/genocideprevention/documents/atrocity-crimes/Doc.1_Convention%20on%20the%20Prevention%20and%20Punishment%20of%20the%20Crime%20of%20Genocide.pdf>

United Nations General Assembly. International Covenant on Civil and Political Rights, 19 December 1966. <https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx>.

United Nations General Assembly. International Covenant on Civil and Political Rights, 19 December 1966. <https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx>.

United Nations Newsroom. "Hate speech on Facebook poses 'acute challenges to human dignity' — UN expert." UN News Human Rights, United Nations, 23 December 2020. <https://news.un.org/en/story/2020/12/1080832>.

United Nations Newsroom. "UN human rights chief points to 'textbook example of ethnic

cleansing' in Myanmar." UN News: Migrants and Refugees, United Nations, 11 Sept 2017. <https://news.un.org/en/story/2017/09/564622-un-human-rights-chief-points-textbook-example-ethnic-cleansing-myanmar>.

United States Holocaust Memorial Museum. "Timeline." Burma's Path to Genocide, USHMM. <https://exhibitions.ushmm.org/burmas-path-to-genocide/timeline>. Accessed 03 Mar 2021.

USA for UNHCR Staff. "Rohingya Refugee Crisis Timeline." USA for UNHCR, United Nations Refugee Agency, 25 Aug 2019. <https://www.unrefugees.org/news/rohingya-refugee-crisis-timeline/>.

Wakabayashi, Daisuke. "Legal Shield for Social Media Is Targeted by Lawmakers." The New York Times, 28 May 2020. <https://www.nytimes.com/2020/05/28/business/section-230-internet-speech.html>.

Waring, Joseph. "Myanmar's MPT Launches Facebook's Free Basics." Mobile World Live, 7 June 2016, <www.mobileworldlive.com/asia/asia-news/myanmars-mpt-launches-facebooks-free-basics.>

Warofka, Alex. "An Independent Assessment of the Human Rights Impact of Facebook in Myanmar." Facebook Newsroom, Facebook, 5 Nov 2018. <https://about.fb.com/news/2018/11/myanmar-hria/>.

Wingfield, Richard. "A Human Rights Approach to Disinformation." Global Partners Digital, 15 Oct 2019. <https://www.gp-digital.org/a-human-rights-based-approach-to-disinformation/>.

Wu, Liang, Fred Morsatter, Kathleen Carley and Huan Liu. "Misinformation in Social Media: Definition, Manipulation, and Detection." *ACM SIGKDD Explorations Newsletter,* Association for Computing Machinery, Nov 2019. <https://doi.org/10.1145/3373464.3373475>.