THE SOCIAL ATROCITY

META AND THE RIGHT TO REMEDY FOR THE ROHINGYA

AMNESTY INTERNATIONAL

Amnesty International is a movement of 10 million people which mobilizes the humanity in everyone and campaigns for change so we can all enjoy our human rights. Our vision is of a world where those in power keep their promises, respect international law and are held to account. We are independent of any government, political ideology, economic interest or religion and are funded mainly by our membership and individual donations. We believe that acting in solidarity and compassion with people everywhere can change our societies for the better.

First published in 2022
by Amnesty International Ltd
Peter Benenson House, 1 Easton Street
London WC1X 0DW, UK

Index: ASA 16/5933/2022
Original language: English

amnesty.org

Cover image: *© Tamara-Jade Kaz*

AMNESTY INTERNATIONAL

CONTENTS

GLOSSARY

WORD	DESCRIPTION
ALGORITHMS	"Algorithmic systems" are understood as applications that, often using mathematical optimisation techniques, perform one or more tasks such as gathering, combining, cleaning, sorting, classifying and inferring data, as well as selection, prioritisation, the making of recommendations and decision making. Relying on one or more algorithms to fulfill their requirements in the settings in which they are applied, algorithmic systems automate activities in a way that allows the creation of adaptive services at scale and in real time.[1]
ARTIFICIAL INTELLIGENCE OR "AI"	There is no widely accepted definition of the term "artificial intelligence" or "AI". The United Nations Office of the High Commissioner for Human Rights uses the term to refer to a constellation of processes and technologies enabling computers to complement or replace specific tasks otherwise performed by humans, such as making decisions and solving problems, including machine learning and deep learning.[2]
CONTENT MODERATION	"Content moderation" refers to social media platforms' oversight and enforcement of platform rules in relation to permissible and prohibited forms of expression. It can include actions such as the detection, demotion and removal of content which violates platform rules.
FACEBOOK PAPERS	A cache of internal Meta documents which were disclosed by whistle-blower Frances Haugen to the US Congress in October 2021.[3]
IIFFMM	The United Nations' Independent International Fact-Finding Mission on Myanmar.
OECD GUIDELINES	OECD Guidelines for Multinational Enterprises
UN GUIDING PRINCIPLES	UN Guiding Principles on Business and Human Rights

[1] Committee of Ministers of the Council of Europe, "Appendix to Recommendation CM/Rec (2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems", 8 April 2020, para. A.2.

[2] UN Office of the High Commissioner for Human Rights (OHCHR), "The right to privacy in the digital age", 15 September 2021, A/HRC/48/31, fn 2.

[3] For further information on the Facebook Papers, see: https://facebookpapers.com/

1. EXECUTIVE SUMMARY

Beginning in August 2017, the Myanmar security forces undertook a brutal campaign of ethnic cleansing against Rohingya Muslims in Myanmar's Northern Rakhine State. They unlawfully killed thousands of Rohingya, including young children; raped and committed other sexual violence against Rohingya women and girls; tortured Rohingya men and boys in detention sites; and burned down hundreds of Rohingya villages. The violence pushed over 700,000 Rohingya – more than 80 per cent of the Rohingya population living in northern Rakhine State at beginning of the crisis – into neighbouring Bangladesh, where most linger in refugee camps to this day. The UN's Independent International Fact-Finding Mission on Myanmar (IIFFMM) called for senior military officials to be investigated and prosecuted for war crimes, crimes against humanity, and genocide.

In the months and years leading up to and during the 2017 atrocities, Facebook in Myanmar became an echo chamber of virulent anti-Rohingya content. Actors linked to the Myanmar military and radical Buddhist nationalist groups systematically flooded the Facebook platform with incitement targeting the Rohingya, sowing disinformation regarding an impending Muslim takeover of the country and seeking to portray the Rohingya as sub-human invaders. The mass dissemination of messages that advocated hatred inciting violence and discrimination against the Rohingya, as well as other dehumanizing and discriminatory anti-Rohingya content, poured fuel on the fire of long-standing discrimination and substantially increased the risk of an outbreak of mass violence.

The IIFFMM concluded that "[t]he role of social media [was] significant" in the atrocities, in a context where the rapid dominance of Facebook's platform in the country had meant that "Facebook is the Internet". The Mission recommended that "[t]he extent to which Facebook posts and messages have led to real-world discrimination and violence must be independently and thoroughly examined." In line with this call, this report provides a first-of-its kind, in-depth human rights analysis of the role played by Meta Systems Inc. (then named Facebook Inc.) in the atrocities perpetrated against the Rohingya in 2017, and Meta's ongoing responsibility to provide remedy to the Rohingya communities.

Sawyeddollah, a 21-year-old Rohingya youth activist, survivor, and now refugee residing in a refugee camp in Bangladesh, reflected on his community's efforts to seek an effective remedy from Meta following the company's prominent role in the atrocities perpetrated against his community:

> **“I really believe that we deserve a remedy from Facebook. I believe that we are living in this hell because of many different [actors], including Facebook… Facebook cannot remake our lives as before; only we can do that. But what we need is education to do it. Facebook has billions of dollars. We are just asking for a little to give opportunities to Rohingya students in these [refugee] camps.”**

This report outlines in detail how Meta – through its dangerous algorithms and its relentless pursuit of profit – substantially contributed to the serious human rights violations perpetrated against the Rohingya. It reveals that Meta’s contribution was not merely that of a passive and neutral platform that responded inadequately in the face of an unprecedented crisis. In reality, Meta’s content-shaping algorithms proactively amplified and promoted content on the Facebook platform which incited violence, hatred, and discrimination against the Rohingya. Ultimately, this is because Meta’s business model, based on targeted advertising itself, fuels the spread of harmful content, including incitement to violence. The algorithmic systems that shape a user’s experience on Facebook and determine what information they see are designed to keep people on the platform – the more engaged users are, the more advertising revenue Meta earns. As a result, these systems prioritize the most inflammatory, divisive and harmful content as it is more likely to “maximize engagement”.

In a typical example of the anti-Rohingya content which proliferated on Facebook at the time, one Facebook post criticizing a human rights defender for his alleged cooperation with the IIFFMM labelled the individual a “national traitor” and consistently added the adjective “Muslim”. The post was shared over 1,000 times and sparked numerous comments calling for the person to be killed, including: “Muslims are dogs and need to be shot.” “Don't leave him alive. Remove his whole race. Time is ticking.” The IIFFMM itself repeatedly tried to report the post to Meta, subsequently describing the company’s response as “slow and ineffective”.

This incitement to hatred, violence and discrimination went to the very top of Myanmar’s military leadership. The leader of Myanmar’s military, Senior General Min Aung Hlaing, posted on his Facebook page on 1 September 2017 saying, “We openly declare that absolutely, our country has no Rohingya race”. Meta finally banned Min Aung Hlaing from Facebook in 2018. He is now leading the country after seizing power in a coup in February 2021.

As the prospect of mass violence against the Rohingya in Myanmar grew, local civil society activists repeatedly pleaded with Meta to act. This report documents in detail the repeated communications and interventions received by Meta between 2012 and 2017, including multiple trips by activists to Meta’s Menlo Park headquarters in California, USA, during which the company was explicitly warned that it risked contributing to a genocide. Despite these efforts, Meta failed to heed the warnings. Inside Myanmar, many Rohingya tried to report anti-Rohingya content via Facebook’s ‘report’ function. However, the company repeatedly failed to enforce its own content policies on removing “hate speech”, allowing these hateful narratives to proliferate and reach unprecedented audiences in Myanmar.

Meta’s wholly inadequate staffing of its Myanmar operations prior to 2017 was a significant factor in the company’s staggering failures to remove harmful anti-Rohingya content from the Facebook platform. This is symptomatic of the company’s broader failure to adequately invest in content moderation across the Global South. As of April 2018, Meta had only five Burmese speakers to monitor and moderate content for Myanmar’s 18 million Facebook users, none of whom were based in Myanmar. Even this small number of staff marked a significant improvement on the situation as it stood prior to the 2017 crisis. In mid-2014, Meta staff admitted that they only had one single Burmese-speaking content moderator devoted to Myanmar at the time, based in their Dublin office.

The risk that Meta could contribute to mass violence against Muslims in Myanmar should have been crystal clear to the company long before the 2017 atrocities perpetrated against the Rohingya. In July 2014, a viral Facebook post triggered an outbreak of lethal violence between Buddhist and Muslim groups in the city of Mandalay. The post falsely claimed that two Muslim men were to blame for the rape of a Buddhist girl in the city. The ensuing riots led the Myanmar authorities to temporarily block Facebook in recognition of the key role played by the platform in fuelling the "instigation" of this violence. Yet Meta's minimal efforts to respond to this overt warning fell dramatically short. In fact, they may have even made matters worse.

In one such example, Meta supported a civil society-led initiative known as 'Panzagar' or 'flower speech' by creating a 'flower speech sticker pack'. Users in Myanmar could post 'Panzagar' stickers on posts containing hate speech as a means of promoting peace and countering hatred. However, civil society activists noticed that the stickers were having serious unintended consequences. According to one activist who was involved in promoting 'Panzagar', Facebook's algorithms counted the stickers as one more way people were enjoying a post. Instead of diminishing the number of people who saw a piece of "hate speech", the stickers had the opposite effect of making the posts even more visible and more popular.

After the IIFFMM highlighted the "significant" role of the Facebook platform in the atrocities perpetrated against the Rohingya, and as public criticism of the company's failures grew increasingly severe, Meta eventually admitted in 2018 that "we weren't doing enough to help prevent our platform from being used to foment division and incite offline violence". In the intervening years, Meta has touted certain improvements in its community engagement and content moderation practices in Myanmar. Yet this report finds that these measures have proven wholly inadequate. This is largely because they focused primarily on improved content moderation (i.e., the detection, demotion and removal of content which violates platform rules). This approach rests on the premise that Meta is a neutral arbiter of content. It therefore fails to tackle the significant role played by the company's own algorithms in proactively amplifying anti-Rohingya content, systems which are central to the company's destructive business model.

Amnesty International's analysis of newly available evidence gleaned from internal Meta documents leaked by whistle-blower Frances Haugen – the "Facebook Papers" – enables a shocking new understanding of the true nature and extent of Meta's contribution to harms suffered by the Rohingya. This evidence shows that the core content-shaping algorithms which power the Facebook platform – including its news feed, ranking, and recommendation features – all actively amplify and distribute content which incites violence and discrimination, and deliver this content directly to the people most likely to act upon such incitement.

As a result, content moderation is inherently inadequate. Internal documents recognize these limitations, with one document from July 2019 stating, "we only take action against approximately 2% of the hate speech on the platform". Another document reveals that some Meta staff, at least, recognize the limitations of content moderation as a solution to algorithmically amplified harms. As one internal memo dated December 2019 reads: "We are never going to remove everything harmful from a communications medium used by so many, but we can at least do the best we can to stop magnifying harmful content by giving it unnatural distribution."

What's more, this report reveals that Meta has long been aware of the risks associated with its algorithms yet failed to act appropriately in response. Internal studies stretching back to as early as 2012 have consistently indicated that Meta's content-shaping algorithms could result in serious real-world harms. In 2016, before the 2017 atrocities in Northern Rakhine State, internal Meta research clearly recognized that "[o]ur recommendation systems grow the problem" of extremism. These internal studies could and should have triggered Meta to implement effective measures to mitigate the human rights risks associated with its algorithms, but the company repeatedly failed to act.

Rather than addressing these risks appropriately, the Facebook Papers reveal in damning detail how Meta continues to ignore such risks to this day in the relentless pursuit of profit. As one former Meta employee outlined in an internal document dated August 2019:

> **"We have evidence from a variety of sources that hate speech, divisive political speech, and misinformation on Facebook and the family of apps are affecting societies around the world. We also have compelling evidence that our core product mechanics, such as virality, recommendations, and optimizing for engagement, are a significant part of why these types of speech flourish on the platform."**

All companies have a responsibility to respect all human rights wherever they operate in the world and throughout their operations. In order to meet this responsibility, companies must engage in ongoing and proactive human rights due diligence processes to identify, prevent, mitigate and account for how they address their impacts on human rights. For technology companies like Meta, due diligence must also include addressing situations in which "business model-driven practices and technology design decisions create or exacerbate human rights risks". Should a company either cause or contribute to an adverse human rights impact, it has a responsibility to provide effective remediation to affected communities.

Amnesty International's analysis shows how Meta's content-shaping algorithms and reckless business practices facilitated and enabled discrimination and violence against the Rohingya. Meta's algorithms directly contributed to harm by amplifying harmful anti-Rohingya content, including advocacy of hatred against the Rohingya. They also indirectly contributed to real-world violence against the Rohingya, including violations of the right to life, the right to be free from torture, and the right to adequate housing, by enabling, facilitating, and incentivizing the actions of the Myanmar military. However, Meta failed to engage in appropriate human rights due diligence in respect of its operations in Myanmar ahead of the 2017 atrocities. This analysis leaves little room for doubt: Meta substantially contributed to adverse human rights impacts suffered by the Rohingya and has a responsibility to provide survivors with an effective remedy.

Meta is currently facing at least three active cases seeking remediation for the Rohingya. Parallel civil legal proceedings were filed against the company in December 2021 in both the United Kingdom and the United States. Refugee groups in Cox's Bazar have also made direct requests to Meta to remediate them by funding a USD $1 million education project in the refugee camps. The USD $1 million requested by the Rohingya to fund their educational initiative represents just 0.002% of Meta's 2021 profits of $46.7 billion. On 10 February 2021, Meta responded to the Rohingya community's latest request with a rejection, stating: "Facebook doesn't directly engage in philanthropic activities."

Meta's presentation of Rohingya communities' pursuit of remedy as a request for charity portrays a deeply flawed understanding of the company's human rights responsibilities. For the Rohingya survivors of the 2017 atrocities in northern Rakhine State, the great majority of whom still live under conditions of extreme deprivation in refugee camps in Cox's Bazar, Bangladesh, the pursuit of justice and remediation is both a matter of principle and one of urgent material need. According to the UN, Rohingya educational needs in 2022 total US$ 70.5 million. Of this amount, just 1.6% has been actually funded. Mohamed Junaid, a 23-year-old Rohingya refugee, lamented the state of educational provision for the Rohingya in Cox's Bazar:

> **"Though there were many restrictions in Myanmar, we could still [go to] school until matriculation at least. But in the camps, our children cannot do anything. We are wasting our lives under tarpaulin."**
>
> Mohamed Junaid, 23-year-old Rohingya refugee

Despite its partial acknowledgement that it played a role in the 2017 violence against the Rohingya, Meta has to date failed to provide an effective remedy to affected Rohingya communities. Following Meta's refusal to fund the community's request for an education project, a group of Rohingya refugees filed a complaint against the company under the OECD Guidelines for Multinational Enterprises via the Irish National Contact Point (NCP). The complaint was transferred to the US NCP in June 2022. As of August 2022, the complaint remained under consideration.

Meta's refusal to compensate Rohingya victims to date – even where the community's modest requests represent crumbs from the table of the company's enormous profits – simply add to the perception that this is a company wholly detached from the reality of its human rights impacts. Far from its headquarters in Menlo Park, Meta's content-shaping algorithms are fanning the flames of hate, violence, and discrimination – and disproportionally impacting the most marginalized and oppressed communities across the world, and particularly in the Global South.

As detailed throughout this report, Meta's flagrant disregard for human rights has proven to be utterly devastating to the Rohingya. Yet these findings are not only relevant to Rohingya survivors; they should sound the alarm that Meta risks contributing to serious human rights abuses again. Already, from Ethiopia to India and other regions affected by conflict and ethnic violence, Meta represents a real and present danger to human rights. Urgent, wide-ranging reforms are needed to ensure that Meta's history with the Rohingya does not repeat itself elsewhere.

And yet, it would be a mistake to conclude that Meta can solve these problems alone. The root cause of Meta's horrendous human rights impacts is hard-wired into the company's business model based on invasive surveillance and profiling – a business model that is shared by other Big Tech companies. Meta's data-hungry business model incentivizes the company to rapidly expand its operations across the globe and into local contexts that entail serious human rights risks, including settings affected by conflict. Big Tech has proven itself incapable of addressing these issues in the absence of effective state regulation. It is imperative that states fulfil their obligation to protect human rights by introducing and enforcing effective legislation to rein in surveillance-based business models across the technology sector.

For the Rohingya, although the true scale of the losses they have suffered is incalculable, systemic change and effective remediation cannot come soon enough. As 22-year-old Showkutara told Amnesty International:

> **"Facebook must pay. If they do not, we will go to every court in the world. We will never give up in our struggle."**

KEY RECOMMENDATIONS

TO META

Remedy

- Work with survivors and the civil society organizations supporting them to provide an effective remedy to affected Rohingya communities.
- Cooperate fully with the OECD NCP process in the United States, and any other processes that may arise from this complaint, and fully fund the education programming requested by Rohingya communities who are parties to the complaint.

Human rights due diligence

- Undertake a comprehensive review and overhaul of human rights due diligence at Meta, including by mainstreaming human rights considerations throughout all Meta platforms' operations, and ensuring that due diligence addresses the systemic and widespread human rights impacts of Meta's business model as a whole.
- Undertake constant, ongoing and proactive human rights due diligence throughout the lifecycle of algorithmic technologies, so that risks and abuses can be identified during the development stage but also after such technologies have been launched.

Business model and algorithms

- Cease the collection of invasive personal data which undermines the right to privacy and threatens a range of human rights.
- End the practice of using tracking-based advertising and embrace less harmful alternative business models, such as contextual advertising.
- Ensure content-shaping algorithms are not based on profiling by default and require an opt-in instead of an opt-out, with the consent for opting in being freely given, specific, informed and unambiguous.
- Radically improve transparency in relation to the use of content-shaping and content moderation algorithms, ensuring that their mechanics are publicly available in clearly understandable terms.

Global South

- Ensure appropriate investment in local-language resourcing throughout the world, with a particular emphasis on resolving existing inequalities that disproportionately impact Global South countries.

TO META'S 'HOME' STATES INCLUDING USA AND IRELAND, AND REGIONAL BODIES SUCH AS THE EU

- Ban targeted advertising on the basis of invasive tracking practices, such as cross-site tracking, and tracking based on sensitive data or other personal data.
- Introduce obligations for platform companies to ensure they address systemic risks to human rights stemming from the functioning and use made of their services.
- Legally require companies, including social media companies, to conduct human rights due diligence and report publicly on their due diligence policies and practices.
- Regulate technology companies to ensure that content-shaping algorithms used by online platforms are not based on profiling by default and must require an opt-in instead of an opt-out, with the consent for opting in being freely given, specific, informed and unambiguous.

2. METHODOLOGY

This report is based on research conducted by Amnesty International between February and June 2022. Amnesty International conducted remote interviews with ten Rohingya activists and refugees, predominantly located in Cox's Bazar, Bangladesh. The organization conducted 12 further interviews with subject matter experts, including former Meta staff, civil society activists involved in digital rights activism in Myanmar, academics, and lawyers.

Most interviews were conducted in English, while some were conducted in the Rohingya language with the support of an interpreter. The information gathered from these interviews was then corroborated with local activists, news coverage, journalists, and other available sources. All interviewees gave informed consent in advance of being interviewed. Amnesty International did not provide any incentives in exchange for interviews. Due to security risks, some of those interviewed requested anonymity, while others wished to share their identities publicly. For those who chose anonymity, Amnesty International used pseudonyms and omitted all potentially identifying information from this report.

Amnesty International also conducted extensive analysis of the human rights implications of the Facebook Papers, a cache of internal Meta documents which were disclosed by whistle-blower Frances Haugen to the US Congress in October 2021. These documents were obtained from public sources, including those published by Gizmodo[4] and Accountable Tech.[5] Additionally, unpublished files were obtained from journalists who had been granted access to the files. This analysis was informed by Amnesty International's interviews with subject matter experts and desk research. The organization also carried out extensive desk research using information from open sources, including relevant international human rights standards, reports from civil society organizations, domestic and international news media, academic journals, and UN reports.

This report builds directly on extensive previous investigations by Amnesty International that documented institutionalized discrimination and segregation of Rohingya Muslims in Myanmar, and crimes against humanity committed during the ethnic cleansing of the Rohingya population in northern Rakhine State in 2017.

On 20 May 2022, Amnesty International wrote to Meta and asked questions regarding the company's actions in relation to its business activities in Myanmar before and during the 2017 atrocities. Meta responded that it could not provide information concerning the period leading up to 2017 because the company is "currently engaged in litigation proceedings in relation to related matters". Amnesty International again wrote to Meta on 14 June 2022 to inform the company of relevant allegations contained in this report and to give the company the opportunity to respond, but Meta declined.

Throughout this report, the term "Meta" is used to refer to the company, Meta Systems Inc (formerly Facebook Inc), including in relation to the period prior to the company's re-brand in October 2021. The term "Facebook" is generally used to refer to the Facebook social media platform, unless directly citing another source that uses "Facebook" to refer to the company itself.

[4] Dell Cameron and others, "Read the Facebook Papers for Yourself", Gizmodo, 18 April 2022, https://gizmodo.com/facebook-papers-how-to-read-1848702919

[5] Accountable Tech, "The Facebook Papers: SEC Documents", undated, https://facebookpapers.com/sec-documents/

3. BACKGROUND

3.1 THE ROHINGYA IN MYANMAR: A HISTORY OF PERSECUTION

The Rohingya are a predominantly Muslim ethnic minority residing primarily in the northern part of Rakhine State, situated in the west of Myanmar. Rakhine State, home to several ethnic and religious groups, is one of Myanmar's poorest and most under-developed regions. The Rohingya have been subjected to decades of state-sponsored discrimination, persecution, and oppression that has been extensively documented by Amnesty International and others.[6] UN Secretary-General Antonio Guterres has referred to the Rohingya as "one of, if not the, most discriminated people in the world".[7]

The Myanmar authorities severely restrict the Rohingya's freedom of movement, effectively segregating them from the rest of society.[8] Access to health care, education, and work opportunities have also been severely limited.[9] Amnesty International has concluded that this discriminatory and dehumanizing regime, which targets the Rohingya as a racial group and is implemented by the state through a range of laws, policies, and practices, constitutes a widespread and systematic attack on a civilian population, and amounts to the crime against humanity of apartheid.[10] As one Rohingya youth described the situation to Amnesty International: "We were born into oppression, in a prison, in Myanmar. Our fathers, their parents, and their parents only ever knew human rights violations."[11]

The situation of the Rohingya deteriorated significantly from 2012 onwards. Violence erupted in several waves between mainly Buddhist ethnic Rakhine civilians, who were at times supported by the security forces, and Rohingya and other Muslims in Rakhine State.[12] Many were killed, and thousands of homes destroyed, resulting in massive displacement. By 2017, some 127,000 people – mainly Rohingya – remained confined to squalid internally displaced person (IDP) camps and unofficial settlements that they are unable to leave without permission.[13] For Rohingya living outside of the camps, restrictions on their

[6] See, for example, Amnesty International, "*We will destroy everything"*; Amnesty International; Amnesty International, *"Caged without a Roof"*: *Apartheid in Myanmar's Rakhine State* (Index: ASA 16/7484/2017), 21 November 2017; Amnesty International, *"My World Is Finished": Rohingya Targeted by Crimes Against Humanity in Myanmar* (Index: ASA 16/7288/2017), 18 October 2017; "*We Are at Breaking Point": Persecuted in Myanmar, Neglected in Bangladesh* (Index: ASA 16/5362/2016), 19 December 2016; Amnesty International, *The Rohingya Minority: Fundamental Rights Denied* (Index: ASA 16/005/2004), May 2004; Amnesty International, *Rohingya: the Search for Safety* (Index: ASA 13/07/97), September 1997; Amnesty International, *Human Rights Violations against Muslims in the Rakhine* (Index: ASA 16/06/92), May 1992; Human Rights Watch, *"All You Can Do Is Pray": Crimes against Humanity and Ethnic Cleansing of Rohingya Muslims in Burma's Arakan State*, April 2013; Human Rights Watch, *Perilous Plight: Burma's Rohingya Take to the Seas*, May 2009; Human Rights Watch, *The Rohingya Muslims: Ending a Cycle of Exodus?*, September 1996; and the Irish Center for Human Rights, *Crimes against Humanity in Western Burma: The Situation of the Rohingyas*, 2010.

[7] Antonio Guterres, "Opening remarks at press encounter with President of the World Bank, Jim Yong Kim", 02 July 2018, Cox's Bazar, un.org/sg/en/content/sg/speeches/2018-07-02/remarks-press-encounter-world-bank-president-jim-kim

[8] Amnesty International, *"Caged without a Roof"*, pp. 42-58.

[9] Amnesty International, *"Caged without a Roof"*, pp. 59-79.

[10] Amnesty International, *"Caged without a Roof"*, pp. 88-98.

[11] Interview by video call with Sawyeddollah, 6 April 2022.

[12] See, for example, Human Rights Watch, *"All You Can Do Is Pray"*; and Human Rights Watch, *"The Government Could Have Stopped This": Sectarian Violence and Ensuing Abuses in Burma's Arakan State*, August 2012, cited in Amnesty International, "*We will destroy everything"*.

[13] Amnesty International, "*Caged without a Roof"*, pp. 22, 53; UN Office for the Coordination of Humanitarian Affairs (OCHA), Myanmar: IDP Sites in Rakhine State (as of 30 April 2018), reliefweb.int/sites/reliefweb.int/files/resources/Myanmar_IDP_Site_Rakhine_Apr2018.pdf

movement tightened, leaving many communities confined to their villages, struggling to access hospitals, schools, and places they rely on for their livelihoods.[14]

Discrimination against the Rohingya takes place in a wider context of worsening anti-Muslim sentiment and religious intolerance across Myanmar over the past decade, which has at times led to attacks on Muslim communities causing deaths, injuries, and destruction of property.[15] In some parts of Myanmar, local communities have – with the knowledge and at times support of local authorities – established "Muslim free" villages.[16]

Hostility against Muslims has often been fuelled by radical Buddhist nationalist groups such as the Ma Ba Tha (the Myanmar language acronym for The Association for the Protection of Race and Religion) that promote discriminatory ultra-nationalist agendas. Ma Ba Tha's public sentiments often amount to advocacy of hatred constituting incitement to discrimination, hostility, or violence, which, under international human rights law, should be prohibited.[17]

The National League for Democracy (NLD)-led government, which came to power after decades of military rule in 2015, also actively propagated and inflamed anti-Muslim and anti-Rohingya sentiment. This was especially the case after attacks were carried out by the Arakan Rohingya Salvation Army (ARSA) in October 2016 and August 2017. On 26 November 2016, while military operations were still ongoing in northern Rakhine State, state media published an opinion piece that described the Rohingya as "extremists, terrorists, ultra-opportunists and aggressive criminals" as "human fleas" who are "loathed for their stench and for sucking our blood".[18] After the 25 August 2017 attacks (see below), the State Counsellor's official Facebook page regularly posted graphic photographs of Hindus, ethnic Mro and Rakhine villagers allegedly killed by "extremist Bengali terrorists", using a common slur ("Bengali") for the Rohingya which seeks to portray them as migrants from Bangladesh.[19]

3.2 THE 2017 "CLEARANCE OPERATIONS"

Against a backdrop of decades of systemic discrimination and apartheid perpetrated against the Rohingya population by the Myanmar authorities, on 25 August 2017, a Rohingya armed group known as the Arakan Rohingya Salvation Army (ARSA) attacked a range of security and military targets in northern Rakhine State. Amnesty International found that ARSA fighters also massacred Hindu civilians on 25 and 26 August in Rakhine State.[20] In the months that followed, the Myanmar security forces, led by the Myanmar Army (also known as Tatmadaw), systematically attacked the entire Rohingya population in villages across northern Rakhine State.[21]

The Myanmar Army would characterize these attacks as "clearance operations", supposedly targeting ARSA insurgents in the region. In the months that followed, more than 702,000 people, including children – over 80 per cent of the Rohingya who lived in northern Rakhine State at the crisis's outset – fled to neighbouring Bangladesh. In Myanmar and elsewhere, at the heart of what is often called "ethnic cleansing," which is not a legal term, is an organized deportation operation – that is, coordinated action aimed at forcing people to leave their homes and country and at ensuring they do not return.[22] The ethnic cleansing of the Rohingya population was achieved by a relentless and systematic campaign in which the Myanmar security forces unlawfully killed thousands of Rohingya, including young children; raped and committed other sexual violence against hundreds of Rohingya women and girls; tortured Rohingya men and boys in detention sites;

[14] Amnesty International, "*Caged without a Roof*", pp. 42-79.

[15] See, for example, Burma Human Rights Network (BHRN), *Persecution of Muslims in Burma*, September 2017; Physicians for Human Rights, *Massacre in Central Burma: Muslim Students Terrorized and Killed in Meiktila*, May 2013; Tomás Ojea Quintana, Special Rapporteur on the situation of human rights in Myanmar, Report, UN Doc: A/68/397, 23 September 2013, paras 58-61; and Amnesty International, *Myanmar: Investigate violent destruction of mosque buildings*, 24 June 2016.

[16] BHRN, *Persecution of Muslims in Burma* (previously cited), pp. 40-50.

[17] International Covenant on Civil and Political Rights (ICCPR), Article 20: *"Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law."*

[18] The Global New Light of Myanmar, "A Flea Cannot Make a Whirl of Dust, But…," 26 November 2016, cited in Amnesty International, "*We will destroy everything",* p.21.

[19] Amnesty International, "*We will destroy everything*", p. 21.

[20] Amnesty International, "Myanmar: New evidence reveals Rohingya armed group massacred scores in Rakhine State", 22 May 2018, amnesty.org/en/latest/news/2018/05/myanmar-new-evidence-reveals-rohingya-armed-group-massacred-scores-in-rakhine-state

[21] Amnesty International, "*We will destroy everything*", p. 8.

[22] Amnesty International, "*We will destroy everything*", p. 137.

pushed Rohingya communities toward starvation by burning markets and blocking access to farmland; and burned hundreds of Rohingya villages in a targeted and deliberate manner.[23]

The actions of the Myanmar authorities against the Rohingya were often supported by the active participation of local vigilantes in Rakhine State. The attacks were conducted with intense brutality. These attacks are extensively documented in previous reports by Amnesty International.[24] Survivors who made the journey to Bangladesh faced further horrors on their way, with many coming under attack from the Myanmar security forces and Rakhine civilians. Each of the Rohingya refugees interviewed by Amnesty International in the course of this research had their own harrowing escape story to tell. Mohamed,[25] recounted his journey to Amnesty International:

> **"On 2nd September, they came to our village... They looted everything like our motorbikes and other expensive things, then they burned down our houses. So we stayed in the jungle for 15 days, and they kept getting closer day by day, burning every village. We knew if we stayed, they would come to us, so we kept walking village to village until finally we arrived in Bangladesh after 122 days of walking."**

During the journey, Mohamed, aged 20, witnessed both of his grandfathers being killed by Myanmar security forces:

> **"One of my grandfathers was killed, he was 70 years old so he could not walk like us. They killed and burned my grandfather; we never saw his face again… My other grandfather came with us for 12 days, he crossed into no man's land on the border, but there was a station of border guard police there. We held our hands up to surrender in front of them, but they opened fire on us, Saweyddollah Salam, my grandfather, was killed by a bullet to the brain, and the same bullet hit his wife. He was immediately dead."**

Mohamed survived the massacre by concealing himself on the ground:

> **"I lay down on the mud, so I survived. Even after we crossed, we still could not say anything. They fired for 20 minutes and 19 people were killed, six more were injured. After they stopped firing, we moved gradually north to the river until we could get a small boat to Bangladesh… On the way, we saw so many dead people who have been killed – so many dead bodies."26**

Tun, who was 16 years old at the time, recounted the horror of his escape on foot from Rakhine State, when his group was attacked by both local vigilantes and the military:

> **"We were nearly 20,000 people in a group, mostly women and children. It took 41 days to reach Bangladesh. Every day we felt that our lives would be ended here. Every day we heard the gunfire noise. Every day. Some of the Buddhist people and military came when we were walking in our line, and they tried to attack us. The people at the back of our line were attacked with knives and guns. Most of them were women and children, and some elderly people. Over 200-300 were killed. They rounded them up and shot and stabbed them. "[27]**

An in-depth investigation by Amnesty International published in 2018 found that the Myanmar authorities' actions amounted to crimes against humanity under international law, as they were perpetrated as part of a widespread and systematic attack against the Rohingya population.[28] During this 10-month period, Amnesty International found evidence of nine of the 11 crimes against humanity listed in the Rome Statute of the International Criminal Court being committed since 25 August 2017, including murder, torture, deportation or forcible transfer, rape and other sexual violence, persecution, enforced disappearance, and other inhumane acts, such as forced starvation.[29] Amnesty International also found that responsibility for these crimes extends to the highest levels of the military, including Senior General Min Aung Hlaing, who is now leading the country after the military seized power from the civilian government in a coup on 1 February

[23] Amnesty International, "*We will destroy everything*".

[24] Amnesty International, "*We will destroy everything*"; Amnesty International, "*My World is Finished*".

[25] Interview by video call with Mohamed (full name withheld for security reasons), Rohingya refugee, 11 April 2022.

[26] Interview by video call with Mohamed (full name withheld for security reasons), Rohingya refugee, 11 April 2022.

[27] Interview by video call with Tun (pseudonym, real name withheld for security reasons), Rohingya refugee, 11 April 2022.

[28] Amnesty International, "*We will destroy everything*".

[29] Amnesty International, "*We will destroy everything*", p. 8.

2021.[30] Since the coup, Amnesty International has documented further war crimes and crimes against humanity being perpetrated by the Myanmar military against civilians in Kayin and Kayah States.[31]

Amnesty International additionally found that during this period the Myanmar security forces violated human rights on a massive scale, including the rights to life and freedom from torture; and the rights to adequate housing and an adequate standard of living, including the right to food; the right to education, particularly for children; and to non-discrimination, which must be fully respected at all times, including during emergencies and situations of armed conflict.[32]

The UN Human Rights Council established the Independent International Fact-Finding Mission on Myanmar (IIFFMM) in March 2017 to investigate and report on the human rights situation in Kachin, Rakhine and Shan States.[33] By September 2018, according to the IIFFMM, "[i]n much of northern Rakhine State, every trace of the Rohingya, their life and community as it had existed for decades, was removed".[34] For the Rohingya that remained in Myanmar after the 2017 atrocities, persecution and apartheid remain a daily reality, and justice and accountability seem more distant than ever since the February 2021 military coup.[35]

The IIFFMM, following its in-depth investigation into the Myanmar military's 'clearance operations' in 2017, considered whether the crime of genocide had been committed against the Rohingya. It found that "the factors allowing the inference of genocidal intent [were] present".[36] The Mission noted: "The actions of those who orchestrated the attacks on the Rohingya read as a veritable check-list [for genocidal intent]: the systematic stripping of human rights, the dehumanizing narratives and rhetoric, the methodical planning, mass killing, mass displacement, mass fear, overwhelming levels of brutality, combined with the physical destruction of the home of the targeted population, in every sense and on every level."[37] In March 2022, the government of the United States declared that a genocide had been perpetrated against the Rohingya during the 2017 atrocities.[38]

3.3 META'S ENTRY INTO MYANMAR

From 1962 to 2011, Myanmar was ruled by a military junta which enforced extreme censorship over all forms of publication. After 2011, however, a nominally civilian government enacted major reforms to the telecommunications sector with a drastic impact on connectivity. In 2011, mobile penetration was estimated at 2% while internet penetration stood at just 0.23% of Myanmar's population – among the lowest rates in Asia.[39] By 2017, mobile penetration had skyrocketed to 93% and internet penetration to 26%.[40]

Even before internet access became widely available in Myanmar, Facebook was already the dominant platform in the country. In 2011, despite widespread restrictions on internet access, it was estimated that approximately 80% of Myanmar's few internet users had Facebook accounts.[41] By 2014, Facebook was estimated to have had hundreds of thousands of users in Myanmar, many of whom were accessing the platform through internet cafes or devices belonging to their friends or family members.[42] Prior to 2015,

[30] Amnesty International, "*We will destroy everything*", p. 8.

[31] Amnesty International, "*Bullets rained from the sky: War crimes and displacement in Eastern Myanmar*", 31 May 2022, ASA 16/5629/2022, amnesty.org/en/documents/asa16/5629/2022/en

[32] Amnesty International, "*We will destroy everything*", p. 135.

[33] The UN Human Rights Council established the IIFFMM in resolution 34/22 in March 2017. UNOHCHR, "Independent International Fact-Finding Mission on Myanmar", ohchr.org/en/hr-bodies/hrc/myanmar-ffm/index

[34] Independent International Fact-Finding Mission on Myanmar, 'Report of the detailed findings of the Independent International Fact-Finding Mission on Myanmar' (IIFFMM, Detailed findings), 17 September 2018, A/HRC/39/CRP.2, para. 1439.

[35] See: Amnesty International, 'Myanmar: World must act now to prevent another year of intolerable 'death and misery'', 27 January 2022, amnesty.org/en/latest/news/2022/01/myanmar-coup-one-year-anniversary

[36] IIFFMM, "Detailed findings" (previously cited), para. 1441.

[37] IIFFMM, "Detailed findings", para. 1440.

[38] Amnesty International, "Myanmar: Momentum for justice as US to label Rohingya crackdown genocide", 21 March 2022, amnesty.org/en/latest/news/2022/03/us-myanmar-rohingya-genocide

[39] Datareportal, "Digital 2011: Myanmar". 28 December 2011, datareportal.com/reports/digital-2011-myanmar

[40] Datareportal, "Digital 2017: Myanmar", datareportal.com/reports/digital-2017-myanmar

[41] Datareportal, "Digital 2011: Myanmar" (previously cited).

[42] Victoire Rio, 'The Role of Social Media in Fomenting Violence: Myanmar', 2020, Toda Peace Institute, Policy Brief No. 78, toda.org/assets/files/resources/policy-briefs/t-pb-78_victoire-rio_role-of-social-media-in-fomenting-violence-myanmar.pdf ; Datareportal, "Digital 2011: Myanmar" (previously cited).

Facebook was only available in Myanmar via an English-language interface, but a specific Myanmar version was launched in 2015.[43]

Facebook usage exploded alongside internet access after 2014, and by 2016, it was estimated that there were 10 million Facebook users in Myanmar.[44] By 2018, the number was estimated to have reached 20 million.[45] Smartphones generally came with the Facebook app pre-installed, and shop owners would often set up Facebook profiles for their customers.[46] Meta's exponential growth in Myanmar was aided by the July 2016 launch of "Free Basics" and "Facebook Flex" in Myanmar - initiatives jointly provided by Meta and Myanmar Post and Telecommunications – a government agency.[47] Facebook Flex is a product that enables subscribers to have a text-only version of Facebook without incurring data charges. Free Basics provided users with access to a basic version of the Facebook platform along with a limited number of services without incurring data charges on their mobile phones.[48]

A 2019 UN report found that with Meta's Free Basics, "more local data would mean opportunities for providing better targeted advertising".[49] Amnesty International has previously found that Free Basics acts as a means for Meta to collect masses of data from people in the Global South. Although Free Basics is presented by Meta as a philanthropic initiative providing an "onramp to the broader internet" for those in the Global South who would otherwise lack internet access, Free Basics instead appears to be an "onramp" for increasing data mining in the Global South.[50]

Under decades of military rule, the population of Myanmar was denied access to diverse media and news sources, and opportunities to express their ideas and opinions were severely curtailed. The rapid transformation of Myanmar's telecommunications landscape and the sudden dominance of the Facebook platform occurred in a context in which digital media literacy was extremely low. As an American platform populated by trusted friends and family, Facebook was widely perceived as a reliable source of news and information. The IIFFMM observed that "the Government's use of Facebook for official announcements and sharing of information further contributes to users' perception of Facebook as a reliable source of information".[51]

Meta's rapid market entry into Myanmar, combined with the enthusiastic embrace of the platform by a population which had long been starved of space to freely express themselves, led to Meta enjoying near-total market dominance in Myanmar by 2017. The Facebook platform was used not only as a means of communicating with friends, but for many people in Myanmar, it became their primary news source, business directory, online marketplace, and go-to search engine.[52] As noted by the IIFFMM in 2018, "[f]or many people, Facebook is the main, if not only, platform for online news and for using the Internet more broadly".[53]

As outlined in Section 3.1, above, Meta's rise to dominance in Myanmar occurred against a backdrop of rising inter-communal tensions and increasing levels of conflict in Northern Rakhine State. At the national level, Myanmar was in the throes of a fragile transition towards democracy after decades of military rule.[54] A number of ethnic minority groups had historical and ongoing conflicts with the military, including in Kayin (Karen) State, Kayah (Karenni) State, Kachin State, and Mon State.[55] As previously noted, the Rohingya in Rakhine State, in particular, were living under a system of apartheid. This context amounted to a conflict-affected setting, as discussed further in Section 4.1, below.

[43] IIFFMM, "Detailed findings", para. 1344.

[44] Catherine Trautwein, "Facebook racks up 10m Myanmar users", 13 June 2016, The Myanmar Times, mmtimes.com/business/technology/20816-facebook-racks-up-10m-myanmar-users.html

[45] IIFFMM, Detailed findings, para. 1344.

[46] See: Rio, "The Role of Social Media in Fomenting Violence" (previously cited); IIFFMM, Detailed findings, para. 1344.

[47] IIFFMM, "Detailed findings", para. 1344.

[48] IIFFMM, "Detailed findings", para. 1344.

[49] UNCTD, Digital Economy Report 2019 (previously cited), p 90.

[50] Amnesty International, "*Surveillance Giants*" (previously cited), p.14.

[51] IIFFMM, "Detailed findings", para. 1345.

[52] IIFFMM, Detailed findings, para. 1345.

[53] IIFFMM, "Detailed findings", para. 1345.

[54] See, Amnesty International, "Annual Report 2013: The State of the World's Human Rights: Myanmar", POL 10/001/2013, amnesty.org/en/documents/pol10/001/2013/en

[55] See: Amnesty International, "Annual Report 2012: The State of the World's Human Rights: Myanmar", POL 10/001/2012, amnesty.org/en/documents/pol10/001/2012/en

4. LEGAL FRAMEWORK

This chapter outlines the international human rights laws and standards which are most relevant to Meta's operations in Myanmar. The chapter begins by outlining international human rights standards as they pertain to corporate actors. It then considers the application of these standards to the use of algorithmic technologies by technology companies. It then surveys international human rights law and standards regarding the right to remedy for individuals and communities whose rights have been impacted by corporate actors. Finally, this chapter proves an overview of international human rights law and standards pertaining to "hate speech" and advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence (advocacy of hatred).

4.1 BUSINESS AND HUMAN RIGHTS STANDARDS

All companies have a responsibility to respect all human rights wherever they operate in the world and throughout their operations. This is a widely recognized standard of expected conduct as set out in international business and human rights standards, including the UN Guiding Principles on Business and Human Rights (UN Guiding Principles) and the OECD Guidelines for Multinational Enterprises (OECD Guidelines).[56]

The corporate responsibility to respect human rights is independent of a State's own human rights obligations and exists over and above compliance with national laws and regulations protecting human rights.[57] UN Guiding Principle 13 outlines that companies' responsibility to respect human rights entails a requirement to "avoid causing or contributing to adverse human rights impacts through their own activities, and address such impacts when they occur; and seek to prevent or mitigate adverse human rights impacts that are directly linked to their operations, products or services by their business relationships, even if they have not contributed to those impacts." [58]

The UN Guiding Principles establish that to meet their corporate responsibility to respect human rights, companies should have in place ongoing and proactive human rights due diligence processes to identify, prevent, mitigate and account for how they address their impacts on human rights. When conducting human rights due diligence, a company may identify that it may cause or contribute to – or already be causing or contributing to – a serious human rights abuse through its own activities. In these cases, companies must cease or prevent the activities that are responsible for those adverse human rights impacts.[59] Where impacts are outside of the business enterprise's control but are directly linked to their operations, products or services through their business relationships, the UN Guiding Principles require the company to seek to mitigate the human rights impact by exercising leverage, or seek to improve leverage where leverage is

[56] This responsibility was expressly recognized by the UN Human Rights Council on 16 June 2011, when it endorsed the UN Guiding Principles on Business and Human Rights (UN Guiding Principles), and on 25 May 2011, when the 42 governments that had then adhered to the Declaration on International Investment and Multinational Enterprises of the OECD unanimously endorsed a revised version of the OECD Guidelines for Multinational Enterprises. See Human Rights and Transnational Corporations and other Business Enterprises, Human Rights Council, Resolution 17/4, UN Doc A/HRC/RES/17/4, 6 July 2011; OECD Guidelines for Multinational Enterprises, OECD, 2011, oecd.org/ corporate/mne

[57] UN Guiding Principles, Principle 11 including Commentary.

[58] UN Office of the High Commissioner for Human Rights (UNOHCHR), 'Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework', 2011, UN Doc HR/PUB/11/04, Principle 13 including Commentary, ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf

[59] UNOHCHR, Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework (2011), ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf

limited, including through collaboration if appropriate.

Principles 17 to 20 of the UN Guiding Principles describe the human rights due diligence responsibilities of corporate actors. Principle 17 states that “the process should include assessing actual and potential human rights impacts, integrating and acting upon the findings, tracking responses, and communicating how impacts are addressed”, and the process “should cover adverse human rights impacts that the business enterprise may cause or contribute to through its own activities, or which may be directly linked to its operations, products or services by its business relationships”. Due diligence practices “[w]ill vary in complexity with the size of the business enterprise, the risk of severe human rights impacts, and the nature and context of its operations” and they should be “ongoing, recognizing that the human rights risks may change over time as the business enterprise’s operations and operating context evolve”.[60]

Principle 18 states that “business enterprises should identify and assess any actual or potential adverse human rights impacts with which they may be involved” including by drawing on “internal and/or independent external human rights expertise” and conducting “meaningful consultation with potentially affected groups and other relevant stakeholders, as appropriate to the size of the business enterprise and the nature and context of the operation”.[61]

Transparency is a key component of human rights due diligence. As the UN Guiding Principles make clear, companies “need to know and show that they respect human rights”[62] and “showing involves communication, providing a measure of transparency and accountability to individuals or groups who may be impacted and to other relevant stakeholders.”[63]

The OECD Guidelines also state that companies should carry out “risk-based due diligence” and state that due diligence processes should seek to “identify, prevent and mitigate actual and potential adverse impacts”.[64] The OECD Guidelines also state that “[e]nterprises should carry out human rights due diligence as appropriate to their size, the nature and context of operations and the severity of the risks of adverse human rights impacts.”[65]

Corporations are subject to higher than usual standards of due diligence in conflict-affected-settings. UN Guiding Principle 23 notes that having operations in conflict-affected areas may increase the risk of being complicit in gross human rights abuses committed by other actors (for example, security forces), which necessitates extra care.[66] The Guiding Principles imply that such measures should take the form of “enhanced” or “heightened” human rights due diligence.[67] A report published in 2020 by the UN Working Group on Business and Human Rights provides additional guidance for businesses operating in conflict-prone regions, and states that "heightened human rights due diligence" should incorporate conflict sensitivity and atrocity-prevention approaches in order to account for the two-way interaction between the business activities and the context, as well as broader stakeholder engagement that includes engagement with armed non-state actors so as to mitigate the information gaps, polarization, and mistrust which usually exists among groups and communities who are in conflict.[68] Notably, the Working Group establishes in this report that, when operating in conflict-affected and post-conflict areas, the business responsibility to remediate human rights harm should also include engagement with transitional justice processes, such as: prosecution initiatives, truth-seeking processes, reparations programmes, and institutional reform.[69]

The report also points to the responsibilities of technology companies, specifically, stating: "There is no exceptionalism [for heightened human rights due diligence in conflict affected settings for the technology] sector” and “the sector should adopt a genuine human rights approach, in which all rights are recognized as equal, rather than the misguided understanding of human rights whereby the right to free speech, or the

[60] UN Guiding Principles, Principle 17.

[61] UN Guiding Principles, Principle 18.

[62] UN Guiding Principles, Commentary to Principle 15.

[63] UN Guiding Principles, Commentary to Principle 21.

[64] OECD Guidelines for Multinational Enterprises, Chapter II ‘General Principles’. 2A, 10.

[65] OECD Guidelines for Multinational Enterprises, Chapter IV ‘Human Rights’, 5.

[66] UN Guiding Principles, Principle 23.

[67] Special Representative of the Secretary-General on the issue of human rights and transnational corporations and other business enterprises, John Ruggie, ‘Business and human rights in conflict-affected regions: challenges and options towards State responses’, 27 May 2011, A/ HRC/17/32, para. 16 (d).

[68] Working Group on the issue of human rights and transnational corporations and other business enterprises, “Business, human rights and conflict-affected regions: towards heightened action”, 21 July 2020, A/75/212.

[69] Working Group on the issue of human rights and transnational corporations and other business enterprises, “Business, human rights and conflict-affected regions: towards heightened action”, 21 July 2020, A/75/212.

right to physical security, would be so absolute or unyielding as to trump any other human rights".[70]

4.2 SOCIAL MEDIA ALGORITHMS AND HUMAN RIGHTS DUE DILIGENCE

The international human rights standards governing responsible business conduct were formulated before the human rights implications of artificial intelligence were widely appreciated – and this has created challenges for those seeking to hold Big Tech companies accountable for their human rights impacts. This accountability gap is exacerbated by the fact that Big Tech's use of algorithms and artificial intelligence has been marked by an absence of transparency regarding how these systems work, and because the inherent complexity of such technologies can require specialized knowledge to unpack and analyse.

WHAT IS A SOCIAL MEDIA ALGORITHM?

Social media algorithms are pieces of computer code comprising a set of rules, which enable automated decisions and recommendations to be made on a social media platform. These algorithmic systems shape a user's experience on the platform and determine what information they see - for example, what post someone sees at the top of their Facebook News Feed. Social media platforms use a wide range of algorithms in the delivery of their services.[71] These algorithms rely on artificial intelligence and machine learning technology, meaning that they operate under varying levels of human oversight. There are two major categories of algorithms which are particularly relevant to this report:

1. **CONTENT-SHAPING ALGORITHMS**

Content-shaping algorithms are designed to serve a user 'relevant' content (relevancy that is inferred by the companies on the basis of collected personal data). These algorithms are continuously trained on vast amounts of user data to serve many different purposes, such as ad targeting and delivery, serving search results, recommending new content, and prompting users to create new content and engage with existing content. To do this, the systems "optimize" to best deliver a specific outcome using highly complex and iterative algorithmic processes that draw correlations and inferences from user data.

Examples of these content-shaping algorithms include the Facebook News Feed and ranking algorithms, which decide what users see on their news feed and in what order, and its recommender algorithms, which decide what content is recommended and auto-played for specific users.

2. **CONTENT MODERATION ALGORITHMS**

Content moderation algorithms are increasingly used to automate social media platforms' detection, demotion, and deletion of content which violates platform rules.[72] The algorithms are partly trained by human staff, who classify and label millions of pieces of content in order to train the algorithms on how to automatically detect violating content. Although content moderation was previously conducted primarily by human moderators, it is increasingly automated. Content moderation algorithms can also take automated decisions in relation to content, including by 'demoting' content in the news feed rankings, or by removing content entirely.

In recent years, international standards have begun to evolve to accommodate these innovations and industries. The UN Office of the High Commissioner for Human Rights (OHCHR)'s B-Tech Project has affirmed that tech companies' due diligence must also include addressing situations in which "business model-driven practices and design decisions create or exacerbate human rights risks", as well as an analysis

[70] Working Group on the issue of human rights and transnational corporations and other business enterprises, "Business, human rights and conflict-affected regions: towards heightened action", 21 July 2020, A/75/212, para. 99.

[71] In 2019, New America's Open Technology Institute published a series of reports setting out how automated tools are being used by internet platforms to shape the content we see and influence how this content is delivered to us. See Spandana Singh, Open Technology Institute, *Everything in Moderation: An Analysis of How Internet Platforms Are Using Artificial Intelligence to Moderate User-Generated Content*, 2019 newamerica.org/oti/reports/everything-moderation-analysis-how-internet-platforms-are-using-artificial-intelligence-moderate-user-generated-content

[72] 'Content moderation' refers to social media platforms' oversight and enforcement of platform rules in relation to permissible and prohibited forms of expression - for Meta, these rules are known as 'Community Standards' (see Section 5.4 below). For additional information, see: "A Tale of Two Algorithms" in Ranking Digital Rights, *It's Not Just the Content, It's the Business Model: Democracy's Online Speech Challenge,* 17 Mach 2020, available at newamerica.org/oti/reports/its-not-just-content-its-business-model

that looks at the unique human rights risks posed by different products and services, end users, and use contexts.[73] In the case of the Facebook platform, Meta should be conducting human rights due diligence on each phase of the product and service lifecycle for their algorithmic technologies – design, development, promotion, sale/licensing, contracting, and use – including in respect of its news feed, ranking, and various recommender algorithms.

Content-shaping algorithms can lead to the creation of social media 'echo chambers', whereby individuals are only shown information which reinforces their pre-existing views and beliefs. The UN Special Rapporteur on the right to freedom of expression warned in 2018 that: "AI [Artificial Intelligence]-driven personalization may also minimize exposure to diverse views, interfering with individual agency to seek and share ideas and opinions across ideological, political or societal divisions. Such personalization may reinforce biases and incentivize the promotion and recommendation of inflammatory content or disinformation in order to sustain users' online engagement".[74]

The Special Rapporteur set out a range of 'Substantive standards for artificial intelligence systems', including the following:

> **"Companies should orient their standards, rules and system design around universal human rights principles (A/HRC/38/35, paras. 41–43). Public-facing terms and guidelines should be complemented by internal policy commitments to mainstreaming human rights considerations throughout a company's operations, especially in relation to the development and deployment of AI and algorithmic systems."[75]**

In 2020, the Council of Europe recommended that states regulate algorithmic systems in order to ensure their human rights compliance, stating:

> **"[E]nsure, through appropriate legislative, regulatory and supervisory frameworks related to algorithmic systems, that private sector actors engaged in the design, development and ongoing deployment of such systems comply with the applicable laws and fulfil their responsibilities to respect human rights in line with the UN Guiding Principles on Business and Human Rights."[76]**

In a 2021 report on disinformation, the UN Special Rapporteur on the right to freedom of expression specifically linked these algorithmic harms to surveillance-based business models in the technology sector, stating: "False information is amplified by algorithms and business models that are designed to promote sensational content that keep users engaged on platforms."[77] The Special Rapporteur also emphasized the limitation of content-based solutions to algorithmically amplified harms, stating:

> **"Reactive content moderation efforts are simply not enough to make a meaningful difference in the absence of a serious review of the business model that underpins much of the drivers of disinformation and misinformation."[78]**

In a 2021 report,[79] the UN Office of the High Commissioner for Human Rights (OHCHR) set out recommendations for addressing human rights risks related to the use of artificial intelligence, including:

a) Systematically conduct human rights due diligence throughout the life cycle of the AI systems they design, develop, deploy, sell, obtain or operate. A key element of their human rights due diligence should be regular, comprehensive human rights impact assessments;

b) Dramatically increase the transparency of their use of AI, including by adequately informing the public and affected individuals and enabling independent and external auditing of automated systems. The more likely and serious the potential or actual human rights impacts linked to the use of AI are, the more transparency is needed;

[73] UNOHCHR, 'Addressing Business Model Related Human Rights Risks: A B-Tech Foundational Paper', August 2020, ohchr.org/Documents/Issues/Business/B-Tech/B_Tech_Foundational_Paper.pdf, p.3.

[74] Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, David Kaye, 'Report on Artificial Intelligence technologies and implications for freedom of expression and the information environment', 29 August 2018, A/73/348, para. 12.

[75] Special Rapporteur on freedom of expression, Report on Artificial Intelligence (previously cited), para. 48.

[76] Committee of Ministers of the Council of Europe, "Recommendation CM/Rec (2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems", 8 April 2020.

[77] Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Irene Khan, "Disinformation and freedom of opinion and expression", 13 April 2021, A/HRC/47/25, para. 16.

[78] Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Irene Khan, "Disinformation and freedom of opinion and expression", 13 April 2021, A/HRC/47/25, para. 65.

[79] OHCHR, "The right to privacy in the digital age", 15 September 2021, A/HRC/48/31.

c) Ensure participation of all relevant stakeholders in decisions on the development, deployment and use of AI, in particular affected individuals and groups;

d) Advance the explainability of AI-based decisions, including by funding and conducting research towards that goal.[80]

There are increasingly efforts among states and lawmakers to introduce binding regulations to tackle the harms of algorithmic systems and hold powerful technology companies to account. In July 2022, the EU's landmark Digital Service Act (DSA) package was adopted by the EU Parliament. The DSA introduces, as a first of its kind, novel obligations on very large online platforms (VLOPs) and very large online search engines (VLOSEs), including a requirement to assess and mitigate systemic risks that arise from the "design, including algorithmic systems, functioning and use made of their services".[81]

4.3 THE CORPORATE RESPONSIBILITY TO PROVIDE REMEDY

All victims of human rights violations and abuses have a right to an effective remedy. This right lies at the very core of international human rights law. It also stems from a general principle of international law that every breach gives rise to an obligation to provide a remedy. The right to an effective remedy has been recognized under various international and regional human rights treaties and instruments and also as a rule of customary international law.[82]

The UN Special Rapporteur on the right to freedom of expression has affirmed that 'adverse impacts of AI [Artificial Intelligence] systems on human rights must be remediable and remedied by the companies responsible'.[83] Similarly, UN OHCHR has called on states to "[e]nsure that victims of human rights violations and abuses linked to the use of AI systems have access to effective remedies".[84]

Access to an effective remedy is a key pillar of the business and human rights framework. The UN Guiding Principles state that where "business enterprises identify that they have caused or contributed to adverse impacts, they should provide for or cooperate in their remediation through legitimate processes".[85]

The OECD Guidelines also state "Enterprises should provide for or cooperate through legitimate processes in the remediation of adverse human rights impacts where they identify that they have caused or contributed to these impacts".[86]

Companies have a varying degree of responsibility to provide an effective remedy to the victims of human rights harms depending on the nature and extent of their role in any adverse human rights impact. The key question in any assessment as to whether a company has a responsibility to provide a remedy to affected individuals or communities is whether the company:

- *Caused the adverse human rights impact,*
- *Contributed to the adverse human rights impact, or was*
- *Directly linked to the adverse human rights impact.*

If a company either *causes* or *contributes* to an adverse impact, it has a responsibility to provide remediation. Where a company does not reach the threshold of 'contribution', it may still be 'directly linked' to an adverse impact. In such cases, the company is not required to provide remediation,[87] however, the company is still expected to use its leverage to prevent or mitigate the impact.[88]

[80] OHCHR, "The right to privacy in the digital age" (previously cited), para. 60.

[81] Amnesty International, *What the EU's Digital Services Act means for human rights and harmful Big Tech business models*, 7 July 2022, amnesty.eu/news/what-the-eus-digital-services-act-means-for-human-rights-and-harmful-big-tech-business-models

[82] For a thorough overview of the right to remedy under international human rights law, see: Amnesty International, "Injustice Incorporated: Corporate abuses and the human rights to remedy", 2014, POL30/001/2014, Chapter 2.

[83] Special Rapporteur on freedom of expression, Report on Artificial Intelligence (previously cited), para. 60.

[84] OHCHR, "The right to privacy in the digital age" (previously cited), para. 59 (g).

[85] UN Guiding Principles, Principle 22.

[86] OECD Guidelines for Multinational Enterprises, Chapter IV 'Human Rights', 6.

[87] UN Guiding Principles, Commentary to Principle 22.

[88] UN Guiding Principles, Principle 13.

A company is deemed to have *caused* an adverse human rights impact if the company's actions, acting alone, resulted in the impact. A company is deemed to have been *directly linked* to an adverse impact where there is a connection between the company's activities and the adverse impact, but where this connection is not substantial enough to reach the higher threshold of *contribution*.

The OECD's Due Diligence Guidance provide additional guidance for making a 'contribution' assessment in this context:

> An enterprise "contributes to" an impact if its activities, in combination with the activities of other entities cause the impact, or if the activities of the enterprise cause, facilitate or incentivise another entity to cause an adverse impact. Contribution must be substantial, meaning that it does not include minor or trivial contributions.
>
> The substantial nature of the contribution and understanding when the actions of the enterprise may have caused, facilitated, or incentivised another entity to cause an adverse impact may involve the consideration of multiple factors. The following factors can be taken into account:
>
> - the extent to which an enterprise may encourage or motivate an adverse impact by another entity, i.e., the degree to which the activity increased the risk of the impact occurring.
> - the extent to which an enterprise could or should have known about the adverse impact or potential for adverse impact, i.e., the degree of foreseeability.
> - the degree to which any of enterprise's activities actually mitigated the adverse impact or decreased the risk of the impact occurring.[89]

According to the OECD, "[t]he mere existence of a business relationship or activities which create the general conditions in which it is possible for adverse impacts to occur does not necessarily represent a relationship of contribution. The activity in question should substantially increase the risk of adverse impact".[90] Where the activity does not substantially increase the risk of an adverse impact, it is more likely that the activity will fall into the 'directly linked' category.

The Guiding Principles distinguish between "actual" and "potential" human rights impact. Actual impact is one that has occurred or is occurring. Potential impact is one that may occur but has not yet done so.[91] According to Guiding Principle 22, actual impact requires remediation. Potential impacts – or human rights risks – require action to prevent harm or mitigate the risks as far as possible and to the extent to which it may do so. Where some residual impact on human rights is unavoidable, this in turn requires remediation.[92]

4.4 "HATE SPEECH" AND ADVOCACY OF HATRED UNDER INTERNATIONAL HUMAN RIGHTS LAW

There is no universally agreed definition of "hate speech" under international human rights law, and expressions of hatred must be considered in light of both the right to freedom of expression and the rights to equality and non-discrimination. The right to freedom of expression protects many forms of speech, even speech which may be deeply offensive, shocking or disturbing.[93] However, the right to freedom of expression is not absolute and it can be restricted under certain circumstances, including for the protection of the rights of others.

Varying definitions of "hate speech" are used by different institutions and actors, including the IIFFMM and Meta. The IIFFMM defines hate speech as "any expression of violent or discriminatory hatred towards people", encompassing forms of expression that "must be prohibited, those that may be prohibited, and those that must not be prohibited but may require a critical response".[94] Meta's hate speech policy has evolved over time – at the time of writing, the company defines hate speech as: "A direct attack against people – rather than concepts or institutions – on the basis of what we call protected characteristics: race, ethnicity, national origin, disability, religious affiliation, caste, sexual orientation, sex, gender identity and

[89] OECD, "OECD Due Diligence Guidance for Responsible Business Conduct", 2018, p.70.

[90] OECD, "Due Diligence Guidance" (previously cited), 2018, p.70.

[91] UNOHCHR, "The Corporate Responsibility to Respect Human Rights: An Interpretive Guide", 2012, HR/PUB/12/02, p.15.

[92] UNOHCHR, "The Corporate Responsibility to Respect Human Rights: An Interpretive Guide", 2012, HR/PUB/12/02, p.18.

[93] UN Human Rights Committee, General Comment 34, CCPR/C/GC/34, para. 11.

[94] IIFFMM, Detailed findings, para. 1307.

serious disease", adding, "We define attacks as violent or dehumanizing speech, harmful stereotypes, statements of inferiority, expressions of contempt, disgust or dismissal, cursing and calls for exclusion or segregation. We also prohibit the use of harmful stereotypes, which we define as dehumanizing comparisons that have historically been used to attack, intimidate or exclude specific groups, and that are often linked with offline violence."[95]

Article 20 of the International Covenant on Civil and Political Rights (ICCPR) explicitly requires states to prohibit by law advocacy of hatred that constitutes incitement to discrimination, hostility or violence .[96] A similar prohibition is contained in Article 4 of the International Convention on the Elimination of All Forms of Racial Discrimination (ICERD).[97] The direct and public incitement to commit genocide is also prohibited under the Genocide Convention.[98] Similarly, under Article 7 of the Universal Declaration of Human Rights, everyone has the right to be free from incitement to discrimination.[99]

The right to equality and non-discrimination is a critical component of international human rights law, constituting a "basic and general principle relating to the protection of human rights",[100] and individuals who have their right to equality infringed must have access to an effective remedy. The Toronto Declaration is a civil society-led statement based on international human rights law which outlines principles for the application of this essential right to the arena of machine learning and artificial intelligence.[101] The Declaration affirms:

> **"Companies and private sector actors designing and implementing machine learning systems should take action to ensure individuals and groups have access to meaningful, effective remedy and redress. This may include, for example, creating clear, independent, visible processes for redress following adverse individual or societal effects, and designating roles in the entity responsible for the timely remedy of such issues subject to accessible and effective appeal and judicial review."[102]**

The Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence – noting the high threshold for defining restrictions on the right to freedom of expression and for the application of Article 20 of the ICCPR – posits a six-part threshold test to guide states' implementation of this prohibition. The six factors that need to be assessed when determining if an expression amounts to advocacy of hatred are: i) Context, ii) Speaker's position or status, iii) Intent, iv) Content and form, v) Extent of the speech act, and vi) Likelihood, including imminence, of harm.[103]

Advocacy of hatred is more than just the expression of ideas or opinions that are hateful towards members of a particular group. It requires a clear showing of intent to incite others to discriminate, be hostile (experience intense and irrational emotions of opprobrium, enmity and detestation) toward, or commit violence against, the group in question. When certain expression constitutes advocacy of hatred, States have an obligation to prohibit it (though not necessarily to criminalize) through a law that is formulated precisely to allow individuals to modify their behaviour in accordance to it. The law and its application must also comply with the ICCPR's provisions on the right to freedom of expression, and in particular must meet the requirements of necessity and proportionality, in compliance with the three-part test in Article 19(3) of the ICCPR.

As various human rights bodies have pointed out, restricting expression in isolation is an ineffective means to combat discrimination, and therefore effective protection and social inclusion of marginalized groups requires broader interventions from the state and other stakeholders. As proposed by the Rabat Plan of Action, tackling the root causes of intolerance requires a much broader set of policy measures, including education on pluralism and diversity, and policies empowering minorities and indigenous people to exercise their right to freedom of expression.[104]

[95] Meta, Hate Speech Policy, https://transparency.fb.com/de-de/policies/community-standards/hate-speech/

[96] International Covenant on Civil and Political Rights (ICCPR), Article 20.2.

[97] International Convention on the Elimination of All Forms of Racial Discrimination (ICERD), Article 4.

[98] The Convention on the Prevention and Punishment of the Crime of Genocide, art. III (c).

[99] Universal Declaration on Human Rights, un.org/en/about-us/universal-declaration-of-human-rights

[100] United Nations Human Rights Committee, General Comment No. 18, UN Doc. RI/GEN/1/Rev.9 Vol. I (1989), para. 1

[101] Amnesty International and Access Now, "The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems", 17 May 2018, POL 30/8447/2018.

[102] Amnesty International and Access Now, "The Toronto Declaration" (previously cited), para. 53.

[103] Rabat Plan of Action (previously cited), para. 29.

[104] Rabat Plan of Action (previously cited), para. 37.

The Rabat Plan of Action distinguishes between forms of expression that advocate hatred that constitute incitement to violence, hostility or discrimination that must be prohibited; forms of expression that are not criminally punishable but may justify a civil suit; and forms of expression that do not give rise to criminal or civil sanctions, but still raise concerns in terms of tolerance, civility and respect for the convictions of others.[105]

For the purposes of this report, the term "advocacy of hatred" refers to expression that advocates hatred constituting incitement to discrimination, hostility or violence and that must therefore be prohibited by law in accordance with Article 20 of the ICCPR. In addition, the report also addresses the spread of expression that may not reach the threshold of "advocacy of hatred", but still raises concerns in terms of tolerance, civility, and respect for the convictions of others, impacting the right to non-discrimination and equality. Amnesty International has not sought to make individual determinations as to whether specific pieces of content on the Facebook platform qualify as "advocacy of hatred". However, as various international human rights bodies have already determined, many of the examples cited in the following chapters would meet this high threshold.

[105] Rabat Plan of Action, para 20.

5. META AND ADVOCACY OF HATRED AGAINST THE ROHINGYA

> **"I started seeing different things on Facebook, some things related to Rohingya people… The writing was in Burmese – it said, 'there is no Rohingya [race] in Myanmar'. I was seeing this post from many people. I thought, what is happening to these people? Why are they posting against us? Why are they against Rohingya people? I started seeing more and more – about 'Bengali' people, that we are not from Myanmar. It said, "if they stay here, we will be under their control". A lot of hate speech, a lot of false news"**
>
> Sawyeddollah, 21, Rohingya activist and survivor[106]

5.1 THE RISE OF SOCIAL MEDIA IN SPREADING ANTI-ROHINGYA SENTIMENT

This section outlines the rising role of social media in relation to the spread of harmful content, including advocacy of hatred, against the Rohingya from 2012 to 2017.

The Facebook platform played an increasingly prominent role in the rise in anti-Muslim and anti-Rohingya sentiment in Myanmar from 2012 onwards. Serious inter-communal violence was sparked in Rakhine State in 2012 following the alleged rape and murder of a 27-year-old Buddhist woman on 28 May 2012, allegedly perpetrated by three Muslim men. Protests and mob violence erupted, and on 3 June, a Yangon-bound bus was stopped by a crowd of some 300 people in Toungup township, and ten Muslim passengers were taken off and beaten to death.[107] In a June 2012 report on the violence, the International Crisis Group, a non-governmental organization, noted the prevalence of "disturbing views posted widely online" and cited "inflammatory rhetoric online" as a possible factor driving the violence, but ultimately concluded that "very low levels of Internet penetration" made it unlikely to be a decisive factor.[108]

[106] Amnesty International interview by video call with Maung Sawyeddollah, 6 April 2022.

[107] International Crisis Group (ICG), Myanmar Conflict Alert: Preventing communal bloodshed and building better relations, 12 June 2012, available at: refworld.org/docid/4fd85cdd2.html [accessed 29 April 2022].

[108] International Crisis Group (ICG), Myanmar Conflict Alert: Preventing communal bloodshed and building better relations, 12 June 2012, available at: refworld.org/docid/4fd85cdd2.html [accessed 29 April 2022].

Each of the Rohingya refugees interviewed by Amnesty International cited a severe deterioration in inter-communal relations from 2012 onwards. Though the Rohingya faced systemic discrimination at the hands of the Myanmar authorities for many decades prior to 2012, they nonetheless reported co-existing relatively harmoniously with other ethnic groups in Rakhine state. Mohamed Ayas, a Rohingya schoolteacher, recalled this marked deterioration to Amnesty International:

> **"We used to live together peacefully alongside the other ethnic groups in Myanmar. Their intentions were good to the Rohingya, but the government was against us. The public used to follow their religious leaders, so when the religious leaders and government started spreading hate speech on Facebook, the minds of the people changed."[109]**

The growth and formalization of radical Buddhist nationalist groups coincided with Meta's rise in Myanmar, and these groups – most infamously Ma Ba Tha – were among the most prominent Facebook users in Myanmar from 2015 onwards.[110] In 2013, U Wirathu, a widely known monk affiliated with Ma Ba Tha, told Time magazine: "[Muslims] are breeding so fast, and they are stealing our women, raping them…They would like to occupy our country, but I won't let them. We must keep Myanmar Buddhist."[111]

Ma Ba Tha monks recognized the opportunity for the Facebook platform to take their anti-Muslim campaign to the next level. In a 2016 interview, U Wirathu told Buzzfeed that, "If the internet had not come to [Myanmar], not many people would know my opinion and messages like now," adding that he had always written books and delivered sermons but that the "*internet is a faster way to spread the messages*". [112]

Serious violence again erupted in July 2014 in Mandalay (the "Mandalay riots") after a rumour was spread online which alleged that two Muslim tea shop owners had raped a Buddhist girl in their employment. The allegations, which later transpired to be false, were shared by U Wirathu on his Facebook page with the comment, "Mafia flame (of the Muslims) is spreading" and that "all Burmans must be ready". Four days of violence ensued, resulting in two people killed and at least 14 reported to have been injured.[113]

Acknowledging the role played by Facebook in fuelling the violence and "instigation", the Myanmar authorities responded by blocking access to the platform in Mandalay.[114] According to a report by Wired, the Myanmar authorities made repeated attempts at contacting Meta staff in order to seek their support in addressing the crisis in Mandalay, and only resorted to blocking Facebook when their efforts at getting in touch failed.[115]

Amnesty International wrote to Meta in May 2022 and asked what steps, if any, the company took after Myanmar authorities blocked the Facebook platform in Mandalay in July 2014, and whether the company assessed the ways in which the platform might be used to amplify advocacy of hatred. Meta responded that "Meta's investments in Myanmar in response to the events in 2017 have been significant", but that the company could not provide information concerning the period leading up to 2017 because the company is "currently engaged in litigation proceedings in relation to related matters".[116]

5.2 ANTI-ROHINGYA "HATE SPEECH" ON FACEBOOK

There are countless examples of anti-Rohingya content shared on Facebook in advance of and during the 2017 atrocities in Northern Rakhine State. This section provides an overview of the extent and nature of this content on the Facebook platform, including examples.

Rohingya survivors of the 2017 atrocities interviewed by Amnesty International told the organization that Facebook was rife with dehumanizing, derogatory, and hateful anti-Rohingya content, much of which likely

[109] Amnesty International interview by video call with Mohamed Showife, 12 April 2022.

[110] Victoire Rio, "The Role of Social Media in Fomenting Violence: Myanmar", 2020.

[111] TIME, "The Face of Buddhist Terror", 1 July 2013, content.time.com/time/subscriber/article/0,33009,2146000,00.html

[112] Buzzfeed News, "This is what happens when millions of people suddenly get the internet", 20 November 2016, buzzfeednews.com/article/sheerafrenkel/fake-news-spreads-trump-around-the-world, cited in Victoire Rio, "The Role of Social Media in Fomenting Violence: Myanmar".

[113] The Diplomat, "The meaning of the Mandalay Riots in Myanmar", 12 July 2014, thediplomat.com/2014/07/the-meaning-of-the-mandalay-riots-in-myanmar

[114] The Diplomat, "The meaning of the Mandalay riots in Myanmar", 12 July 2014, thediplomat.com/2014/07/the-meaning-of-the-mandalay-riots-in-myanmar ; https://burma.irrawaddy.com/opinion/2014/07/04/61420.html

[115] Wired, "How Facebook's rise fueled chaos and confusion in Myanmar", 6 July, 2018, wired.com/story/how-facebooks-rise-fueled-chaos-and-confusion-in-myanmar

[116] Meta letter to Amnesty International, 31 May 2022 (on file with Amnesty International).

amounted to advocacy of hatred, in the months and years leading up to August 2017. Sawyeddollah, a 21-year-old Rohingya refugee and youth activist, explained how he started to see a huge increase in content targeting the Rohingya on the platform after he 'added' individuals from other ethnic groups as Facebook friends:

> **"I started seeing different things on Facebook, some things related to Rohingya people… The writing was in Burmese – it said, 'there is no Rohingya [race] in Myanmar'. I was seeing this post from many people. I thought, what is happening to these people? Why are they posting against us? Why are they against Rohingya people? I started seeing more and more – about 'Bengali' people, that we are not from Myanmar. It said, 'if they stay here, we will be under their control'. A lot of hate speech, a lot of false news."[117]**

A number of UN bodies, NGOs, and media organizations have documented some of the vast quantity of anti-Rohingya content, including content amounting to incitement to violence, discrimination, and genocide, which circulated on Facebook in advance of and during the 2017 atrocities. In the months and years leading up to August 2017, content that spread dehumanizing, hateful and discriminatory views towards the Rohingya – oftentimes portraying genocidal intent – was rife on the Facebook platform throughout Myanmar.[118] This content, which encouraged and justified violence and discrimination against the Rohingya, was posted by a variety of actors, including senior government and military officials, prominent civilian hate groups and figures, including radical Buddhist nationalist groups such as Ma Ba Tha, and 'news' pages, groups, and other accounts with large followings.

The IIFFMM documented "over 150 online public social media accounts, pages and groups [that] regularly spread messages amounting to hate speech against Muslims in general or Rohingya in particular".[119] According to the IIFFMM, these were just a "small sample of the kinds of messages that have circulated in Myanmar in recent years".[120] Many of the examples cited by the IIFFMM contained explicit death threats and promises of violence against the Rohingya, and frequently used dehumanizing language such as *Kway Kalar* ("Muslim dog"), which was also frequently used by the Myanmar military during the 2017 atrocities.[121]

Some of the viral posts cited by the IIFFMM received enormous amounts of engagement. In one such example:

> **"Dr. Tun Lwin, a well-known meteorologist with over 1.5 million followers on Facebook, called on the Myanmar people to be united to secure the "west gate" and to be alert "now that there is a common enemy". He further stated that Myanmar does not tolerate invaders. As of August 2018, the post had 47,000 reactions, over 830 comments, and nearly 10,000 shares. Several comments called for immediate "uprooting" and "eradication" of the Rohingya, citing the situation in Rakhine State as a "Muslim invasion"."[122]**

Over 1,000 posts, comments, images and videos attacking the Rohingya and other Muslims in Myanmar – some of which had been up on the platform since 2012 – were documented and analysed by Reuters news agency and the Human Rights Center at UC Berkeley School of Law in August 2018.[123] One illustrative post cited by the Reuters investigation which had been on Facebook since 2013 stated, "We must fight them the way Hitler did the Jews, damn *kalars*!"; another example cited a post which stated, "These non-human *kalar* dogs, the Bengalis, are killing and destroying our land, our water and our ethnic people."[124]

[117] Amnesty International interview by video call with Maung Sawyeddollah, 6 April 2022

[118] IIFFMM, Reuters, Ray Serrato, Phandeeyar See, for example: UN Human Rights Committee, Report of the detailed findings of the Independent International Fact-Finding Mission on Myanmar, 18 September 2018, UN Doc A/HRC/39/CRP.2; Ray Serrato cited in The Guardian, "Revealed: Facebook hate speech exploded in Myanmar during Rohingya crisis, 3 April 2018, theguardian.com/world/2018/apr/03/revealed-facebook-hate-speech-exploded-in-myanmar-during-rohingya-crisis; Reuters, "Why Facebook is losing the war on hate speech in Myanmar", 15 August 2018, reuters.com/investigates/special-report/myanmar-facebook-hate/

[119] UN Human Rights Council, Report of the detailed findings of the Independent International Fact-Finding Mission in Myanmar, 17 September 2018, UN Doc A/HRC/39/CRP.2, para 131.

[120] UN Human Rights Council, Report of the detailed findings of the Independent International Fact-Finding Mission in Myanmar, 17 September 2018, UN Doc A/HRC/39/CRP.2 , para 1311.

[121] UN Human Rights Council, Report of the detailed findings of the Independent International Fact-Finding Mission in Myanmar, 17 September 2018, UN Doc A/HRC/39/CRP.2 para 1326.

[122] UN Human Rights Council, Report of the detailed findings of the Independent International Fact-Finding Mission in Myanmar, 17 September 2018, UN Doc A/HRC/39/CRP.2 , para 1326.

[123] Reuters, "Why Facebook is losing the war on hate speech in Myanmar", 15 August 2018, reuters.com/investigates/special-report/myanmar-facebook-hate

[124] Reuters, "Why Facebook is losing the war on hate speech in Myanmar", 15 August 2018, reuters.com/investigates/special-report/myanmar-facebook-hate

Further analysis conducted by researcher Raymond Serrato of 15,000 Facebook posts from a Facebook group containing 55,000 Ma Ba Tha supporters from June 2016 onwards found that activity in the group "spiked on 24 and 25 August 2017, when ARSA Rohingya militants attacked government force… with posts registering a 200% increase in interactions" – this being the event that precipitated the launch of the Myanmar military's 'clearance operations'.[125]

The connection between increasingly prevalent anti-Rohingya posts on Facebook and the risk of an outbreak of mass violence has long been noted by NGOs and academics. In September 2017, weeks after the military launched its campaign of violence, the Institute for War and Peace Reporting published findings from its hate speech monitoring project which it had been undertaking during the previous two years in Myanmar:

> **"The vast majority of hate speech was on social media, particularly Facebook… But while not all hate speech was anti-Muslim or anti-Rohingya, the overwhelming majority certainly was… Over time, we saw the hate speech becoming more targeted and militaristic… At the same time we noted… in the absence of any kind of political leadership that a darkening and deepening vacuum that would ultimately result in a violent reckoning."[126]**

5.2.1 THE ACTIVE ROLE OF THE MYANMAR MILITARY AND AUTHORITIES IN SPREADING ANTI-ROHINGYA CONTENT

Amnesty International has previously documented the extensive use of the Facebook platform by senior military officials to justify, encourage, or order violence and discrimination against the Rohingya during and preceding the so-called "clearance operations".[127] The Commander-in-Chief of the Myanmar military, Senior General Min Aung Hlaing had close oversight over the operations, and Amnesty International has documented his involvement in crimes against humanity perpetrated against the Rohingya.[128] On 1 September 2017, Senior General Min Aung Hlaing announced on his Facebook page that, "in the ongoing incidents," the Myanmar military "had to get involved as the strength of police forces alone could not defend."[129] From 19-21 September, while in Rakhine State, posts on Senior General Min Aung Hlaing's Facebook page indicated his active involvement in the "clearance operations", including statements indicating that he "gave [those commanders] instructions on getting timely information" and on the "systematic deployment of security forces," among other things. [130] He also used his Facebook page to justify and celebrate the actions of the military in Northern Rakhine State, stating that he "honoured" the military's "brilliant efforts to restore regional peace."[131]

Amnesty International has documented how Senior General Min Aung Hlaing's Facebook page often included inflammatory posts, including a 1 September 2017 post that said, "We openly declare that absolutely, our country has no Rohingya race'"; and a 20 September 2017 post that said "collective efforts must be made to protect the minorities of Buthidaung/Maungtaw region such as Mro, Khamee, Thet and

[125] The Guardian, "Revealed: Facebook hate speech exploded in Myanmar during Rohingya crisis", 3 April 2018, theguardian.com/world/2018/apr/03/revealed-facebook-hate-speech-exploded-in-myanmar-during-rohingya-crisis

[126] Institute for War and Peace Reporting, "How social media spurred Myanmar's latest violence", 12 September 2017, iwpr.net/global-voices/how-social-media-spurred-myanmars-latest-violence

[127] Amnesty International, *My World Is Finished: Rohingya Targeted In Crimes Against Humanity In Myanmar,* (Index: ASA 16/7288/2017), 18 October 2017, amnesty.org/en/documents/asa16/7288/2017/en; Amnesty International, *"We Will Destroy Everything": Military Responsibility For Crimes Against Humanity In Rakhine State, Myanma*r (Index: ASA/16/8630/2018), 27 June 2018.

[128] Amnesty International, *My World Is Finished: Rohingya Targeted In Crimes Against Humanity In Myanmar,* (Index: ASA 16/7288/2017), 18 October 2017; Amnesty International, *"We Will Destroy Everything": Military Responsibility For Crimes Against Humanity In Rakhine State, Myanma*r (Index: ASA/16/8630/2018), 27 June 2018.

[129] Senior General Min Aung Hlaing Facebook Posts,1 September 2017, facebook.com/seniorgeneralminaunghlaing/posts/1698274643540350 (last accessed 12 October 2017), cited in, Amnesty International, *My world is finished*, p.41, amnesty.org/en/documents/asa16/7288/2017/en

[130] Senior General Min Aung Hlaing Facebook Posts, 19 September 2017, facebook.com/seniorgeneralminaunghlaing/posts/1715340745167073; 20 September 2017, facebook.com/seniorgeneralminaunghlaing/posts/1716465351721279(all last accessed 12 October 2017), cited in, Amnesty International, "*My world is finished"*, p.41.

[131] Senior General Min Aung Hlaing Facebook Posts, 19 September 2017, facebook.com/seniorgeneralminaunghlaing/posts/1715340745167073, 20 September 2017; facebook.com/seniorgeneralminaunghlaing/posts/1716465351721279 (all last accessed 12 October 2017), cited in, Amnesty International, "*My world is finished"*, p.41.

Daingnet"[132] – making no mention of protecting Rohingya civilians. On 11 October, a statement on his Facebook page said it was an "exaggeration to say that the number of Bengalis fleeing to Bangladesh is very large" and that the "native place of the Bengalis is really Bengal. Therefore, they might have fled to the other country with the same language, race and culture as theirs by assuming that they would be safer there."[133]

Lower-ranking soldiers also used Facebook to celebrate their military offensive against the Rohingya, often utilizing dehumanizing and discriminatory language. For example, on 26 August 2017, infantryman Sai Sitt Thway Aung posted to his five thousand Facebook followers stating: "One second, one minute, one hour feels like a world for people who are facing the danger of Muslim dogs."[134]

Facebook was also used by the Myanmar authorities to disseminate disinformation regarding the military's actions in relation to the "clearance operations". In a November 2017 Facebook post, a military investigation team claimed to have found no evidence of wrongdoing by security forces, concluding that "there was no death of innocent people," and "not a single shot was fired" at women and children fleeing their homes.[135] Senior civilian officials in the NLD-led government also frequently posted content on Facebook which sought to justify mass violence against the Rohingya, often in derogatory and dehumanizing terms. After the 25 August 2017 ARSA attacks, the State Counsellor Aung San Suu Kyi's official Facebook page regularly posted graphic photographs of Hindus, ethnic Mro and Rakhine villagers allegedly killed by "extremist Bengali terrorists", using a slur to describe the Rohingya.[136]

An investigation published by The New York Times in October 2018 revealed the role of systematic and long-running military operations in seeking to sow anti-Muslim and anti-Rohingya narratives via the Facebook platform. The investigation found that the Myanmar military "were the prime operatives behind a systematic campaign on Facebook that stretched back half a decade and that targeted the country's mostly Muslim Rohingya minority group".[137] According to the Times' investigation: "[H]undreds of military personnel… created troll accounts and news and celebrity pages on Facebook and then flooded them with incendiary comments and posts timed for peak viewership".[138]

A study by Victoire Rio – a digital rights researcher who has worked extensively with various civil society groups in Myanmar – found that both the Myanmar military and Ma Ba Tha had systematic operations, employing "hundreds" of full-time staff between them, in order to maximize the potential of Facebook to amplify their anti-Muslim and anti-Rohingya sentiments.[139] Underlining the importance of Facebook to senior military figures, Rio claims: "By the end of 2017, the official page of Commander-in-Chief Senior General Min Aung Hlaing had over 1.3 million followers, while his office page counted over 2.7 million".[140]

Michael (pseudonym), an international aid worker who worked with Rohingya communities in Rakhine State from 2013 to 2019, told Amnesty International about how he became the subject of a viral anti-Rohingya posts on Facebook during the 2017 atrocities, and received death threats as a result.

Michael was evacuated from Maungdaw in Rakhine State during the peak of the 2017 violence. After Michael reached Yangon, he gave an interview about the situation in northern Rakhine State to a local journalist, which went viral on Facebook:

> **"The next day that interview had completely blown up – everyone in my office had seen it already, and there was a lot of hate speech directed at me online."**

[132] Senior General Min Aung Hlaing, Facebook Posts,1 September 2017, facebook.com/seniorgeneralminaunghlaing/posts/1698274643540350; and 20 September 2017, facebook.com/seniorgeneralminaunghlaing/posts/1716465351721279(both last accessed 12 October 2017), cited in, Amnesty International, *My world is finished*, p.42.

[133] Senior General Min Aung Hlaing, Facebook Post, 11 October 2017, facebook.com/seniorgeneralminaunghlaing/posts/1736743383026809(last accessed 12 October 2017), cited in Amnesty International, *My world is finished*, p.42.

[134] Cited in Cecilia Kang and Sheera Frenkel, "*An Ugly Truth: Inside Facebook's Battle for Domination*", Chapter 9, p.181.

[135] CINCDS Facebook Post, Information released by the Tatmadaw True News Information Team on the findings of the Investigation Team in connection with the performances of the security troops during the terrorist attacks in Maungtaw region, Rakhine State, 13 November 2017, cited in Amnesty International, "*We will destroy everything*", p.21.

[136] Amnesty International, "*We will destroy everything*", p.21.

[137] Paul Mozur, "A Genocide Incited on Facebook, With Posts From Myanmar's Military", The New York Times, 15 October 2018, nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html

[138] Paul Mozur, "A Genocide Incited on Facebook, With Posts From Myanmar's Military", The New York Times, 15 October 2018, cited above.

[139] See Victoire Rio, *The Role of Social Media in Fomenting Violence: Myanmar*, 2020, Toda Peace Institute, Policy Brief No. 78; Paul Mozur, "A Genocide Incited on Facebook, With Posts From Myanmar's Military", The New York Times, 15 October 2018,cited above.

[140] Victoire Rio, *The Role of Social Media in Fomenting Violence: Myanmar*, 2020, Toda Peace Institute, Policy Brief No. 78

It did not take long before this online vitriol developed into threats against Michael's life:

> **"Later that day, someone sent me a message with a wink saying 'you might want to see this' – it was a picture of my opened passport info page with a diatribe against me – it had already been shared around 200 times when I saw it. People were saying that this guy has to be killed. Get him out of the country. It was specifically on Facebook only that this was circulating."**

Within hours, the risks to Michael's life were seen as so severe that the United Nations warned him that his life was in immediate danger and that he needed to leave the country immediately.[141]

The IIFFMM, in its final report on 12 September 2018, concluded that "[t]he Myanmar authorities, including the Government and the Tatmadaw, have fostered a climate in which hate speech thrives, human rights violations are legitimized, and incitement to discrimination and violence facilitated"; adding, "[t]he role of social media is significant."[142] The IIFFMM also emphasized this was in a context where the rapid dominance of Facebook's platform in the country had meant that "Facebook is the Internet". The IIFFMM report lamented Meta's failure to provide transparent and comprehensive data in relation to advocacy of hatred spread through its platform in Myanmar, and recommended that "[t]he extent to which Facebook posts and messages have led to real-world discrimination and violence must be independently and thoroughly examined."[143]

One month after the IIFFMM highlighted the "significant" role played by the Facebook platform, Meta published an independent human rights impact assessment (HRIA) of its presence in Myanmar conducted by Business for Social Responsibility (BSR).[144] The HRIA acknowledged that Facebook played a role in the violence against the Rohingya in the following terms:

> **"Facebook has become a useful platform for those seeking to incite violence and cause offline harm. Though the actual relationship between content posted on Facebook and offline harm is not fully understood, Facebook has become a means for those seeking to spread hate and cause harm, and posts have been linked to offline violence. A minority of users is seeking to use Facebook as a platform to undermine democracy and incite offline violence, including serious crimes under international law; for example, the Report of the Independent International Fact-Finding Mission on Myanmar describes how Facebook has been used by bad actors to spread anti-Muslim, anti-Rohingya, and anti-activist sentiment."[145]**

Following the publication of the HRIA, a Meta representative acknowledged in respect of Myanmar that "prior to this year, we weren't doing enough to help prevent our platform from being used to foment division and incite offline violence. We agree that we can and should do more."[146] This HRIA is analysed further in Sections 6.3 and 8.3.

As demonstrated by the reports and testimonies cited above, the Facebook platform was awash with discriminatory, dehumanizing, and hateful anti-Rohingya content in the months and years leading up to August 2017, and throughout the 2017 atrocities themselves. This content, which included advocacy of hatred directed against the Rohingya, was inherently harmful from a human rights perspective, resulting in direct impacts on the rights to equality and non-discrimination (as analysed further in Chapter 8, below). The volume and scale of these anti-Rohingya narratives also created a significant risk that this content could result in the enactment of real-world violence. This connection is analysed in the following section.

[141] Amnesty International interview via video call with Michael (pseudonym), 28 March 2022.

[142] IIFFMM, "Report of the independent international fact-finding mission on Myanmar", 12 Sep 2018, A/HRC/39/64 (hereafter "IIFFMM, Final Report"), paras. 73 & 74.

[143] IIFFMM, Final Report, paras. 73 & 74.

[144] BSR, "Human Rights Impact Assessment: Facebook in Myanmar", 2018.

[145] BSR, "Human Rights Impact Assessment: Facebook in Myanmar", 2018, p.24.

[146] Meta, "An Independent Assessment of the Human Rights Impact of Facebook in Myanmar", 5 November 2018, about.fb.com/news/2018/11/myanmar-hria

5.3 ANALYSIS: HOW ANTI-ROHINGYA CONTENT ON FACEBOOK LINKED TO OFFLINE VIOLENCE

As detailed in the preceding section, advocacy of hatred against the Rohingya, as well as other dehumanizing and discriminatory anti-Rohingya content, was rife on the Facebook platform before and during the 2017 atrocities, resulting in direct human rights impacts against the Rohingya. This section provides an analysis of the connection between this online content and the enactment of offline violence.

In its final report, the IIFFMM noted that "[t]he Mission has no doubt that the prevalence of hate speech in Myanmar significantly contributed to increased tension and a climate in which individuals and groups may become more receptive to incitement and calls for violence. This also applies to hate speech on Facebook."[147] The Mission documented a number of specific incidents whereby viral Facebook posts were linked to incidences of violence offline, concluding that "the linkage between offline and online hate speech and real-world acts of discrimination and violence is more than circumstantial".[148]

Academic literature on genocide and other forms of mass violence frequently draws a connection between advocacy of hatred, dehumanization, and acts of mass violence. In relation to dehumanization as a justification for violence, Adam Jones, author of *Genocide: A Comprehensive Introduction*, notes that "[b]efore they are killed, they are brutalized, debased, and dehumanized – turning them into something approaching "subhumans" or "animals" and, by a circular logic, justifying their extermination".[149] Jones underscores the centrality of "the less dramatic, permitted, everyday acts of violence that make participation (under other conditions) in genocidal acts possible", which include, "all expressions of social exclusion, dehumanization, depersonalization, pseudo-speciation, and reification that normalize atrocious behaviour and violence toward others".[150]

Other scholars focus on the function of ideology, and its dissemination, in justifying acts of violence perpetrated in the context of mass atrocities. Jonathan Leader Maynard offers that "Ideology may (a) generate or shape active motives that create the desire to commit violence; (b) create legitimating perceptions or beliefs which make violence seem permissible prior to/during commission; and/or (c) provide rationalizing resources for retrospectively dealing with the commission or permission of violence after the fact".[151]

The phenomenon of "epistemic dependence" is another means of understanding how individuals can be influenced to engage in acts of violence against a target group. It emphasizes the importance of the presence of trusted sources of information, including media sources and civic leaders, which legitimize and promote violence and dehumanization. Maynard states that "atrocity-justifying ideologies are most influential when they operate through such epistemic dependence".[152] Susan Benesch's "Dangerous Speech Project" also highlights the central importance of the medium of dissemination when advocacy of hatred is spread, and specifically cites Facebook in the context of mass violence against the Rohingya as a key example. It notes:

> **"Messages also tend to have a greater capacity to persuade if there are no alternative sources of news available, or if other sources don't seem credible…. As a result, Facebook became a highly influential medium, used to spread frightening, false messages intended to turn the majority population against minority Rohingya Muslims, even as the country's military has carried out a vicious campaign to drive the Rohingya out, including rape, killing, and burning villages."[153]**

Chris Sidoti, an international human rights lawyer and one of the three expert members of the IIFFMM, explained the function of dehumanization of the Rohingya to Amnesty International:

[147] UN Human Rights Council, Report of the detailed findings of the Independent International Fact-finding Mission on Myanmar, 18 September 2018, UN Doc. A/HRC/39/CRP.2, para 1354.

[148]UN Human Rights Council, Report of the detailed findings of the Independent International Fact-finding Mission on Myanmar, 18 September 2018, UN Doc. A/HRC/39/CRP.2, para 2929 .

[149] Adam Jones, *Genocide: A Comprehensive Introduction,* Second Edition, p. 393.

[150] Adam Jones, *Genocide: A Comprehensive Introduction*, Second Edition, p. 437.

[151] Jonathan Leader Maynard, *Rethinking the Role of Ideology in Mass Atrocities,* 25 February 2014, Terrorism and Political Violence, 26:5, 821 -841, DOI: 10.1080/09546553.2013.796934 p.828.

[152] Jonathan Leader Maynard, *Rethinking the Role of Ideology in Mass Atrocities,* 25 February 2014, Terrorism and Political Violence, 26:5, 821 -841, DOI: 10.1080/09546553.2013.79693 p.827.

[153] Dangerous Speech Project, Dangerous Speech: A Practical Guide, 19 April 2021, dangerousspeech.org/guide/

> **"Dehumanization enables humans to undertake anti-human activities by reducing [the victim] to a sub-human level, which enables justification of their killing. It means that setting fire to a house with women and kids inside is nothing more than pouring petrol down an ants' nest. It is an essential step in enabling soldiers – who are human beings – to partake in inhuman acts, and commit atrocities, as we saw in Rakhine state."[154]**

The proposition that the dissemination of advocacy of hatred and dehumanizing narratives leads to violence and discrimination is also supported by empirical studies of previous situations of mass violence. For example, one study of the effects of the propaganda outlet Radio Television Libre des Milles Collines (RTLM) in the context of the Rwandan genocide indicates that killings were 65-77% higher in Rwandan villages that received the RTLM signal, compared with those that did not receive the signal.[155]

In 2014, Myanmar-based academic Matt Schissler raised the alarm about the potential for the spread of anti-Rohingya ideology on Facebook to trigger mass violence in Myanmar, citing risks associated with "the invisible algorithms Google, Facebook and others use to filter content".[156] Schissler warned that Meta and its algorithmic tendency to produce "echo chambers" could become a "productive force that contributes to shaping worldviews and political viewpoints, potentially driving polarization and extremism" in Myanmar.[157] He concluded:

> **"If this paper's analysis is accurate, it would indicate that the expansion of Internet access in Myanmar, particularly the rise of social media, will contribute to dynamics that have the potential to strengthen the ideological justifications necessary for mass violence."[158]**

In a 2020 study retrospectively examining the role of social media in fomenting violence in Myanmar, Victoire Rio emphasized the broad societal impact that Facebook had in setting the ground for an outbreak of mass violence:

> **"Though it is difficult to tie what happened in Northern Rakhine in 2017 to individual pieces of Facebook content, the sustained campaigns waged on Facebook by both the military and Ma Ba Tha from as early as 2012 undoubtedly played a part in creating an enabling environment for the military crackdown and violence that took place. This is evidenced by the fact that the public opinion at the time was overwhelmingly supportive of the military and its use of force, which many within Myanmar saw as being justified on the basis of the narratives they had been fed for years on Facebook."[159]**

As the scholarship on genocide and mass violence contends, the dissemination of ideology is core to the creation of an enabling environment for mass violence. This includes narratives of dehumanization, of impending threat or takeover from the 'other', and false information regarding the wrongs they have supposedly perpetrated. This ideology influences policymakers in their decision-making, directs perpetrators in their individual acts of brutality and murder, and engages 'bystanders' in their support for the violent actions against the target group. In the context of the mass violence perpetrated against the Rohingya, anti-Rohingya ideology played a central role in enabling and justifying the acts of violence perpetrated by both soldiers and civilians in Rakhine State alike.

In respect of soldiers from the Myanmar military, the role of ideology can be most clearly evidenced in the gratuitous violence and brutalization which accompanied the military's core "clearance" mission. The IIFFMM has noted the "extreme brutality of the acts and campaign against the Rohingya" as part of its finding of genocidal intent, noting that the "brutality of an attack is relevant to whether the perpetrators acted with specific intent".[160] The widespread commission of inhumane acts – including rape and severe sexual violence against women and girls, the murder of children, and even the desecration of victims' dead bodies – is relevant evidence both of specific intent to destroy the Rohingya, and also the function of ideology in perpetration of mass violence.

[154] Amnesty International Interview by video call with Chris Sidoti, 13 April 2022.

[155] David Yanagizawa, Propaganda and Conflict: Theory and Evidence from the Rwandan Genocide, 2012, people.su.se/~daya0852/Rwanda_jmp.pdf, cited in Susan Benesch, World Policy Institute, *Dangerous Speech: A Proposal to Prevent Group Violenc*e, 12 January 2012 worldpolicy.org/wp-content/uploads/2016/01/Dangerous-Speech-Guidelines-Benesch-January-2012.pdf

[156] Matt Schissler, "Echo chambers in Myanmar: Social media and the ideological justifications for mass violence", 2014, Paper for the Australian National University Department of Political & Social Change Research Colloquium, "Communal Conflict in Myanmar: Characteristics, Causes, Consequences"

[157] Matt Schissler, "Echo chambers in Myanmar: Social media and the ideological justifications for mass violence" (previously cited), p.4.

[158] Matt Schissler, 'Echo chambers in Myanmar: Social media and the ideological justifications for mass violence" (previously cited).

[159] Victoire Rio, "The Role of Social Media in Fomenting Violence: Myanmar", 2020, Toda Peace Institute, Policy Brief No. 78.

[160]IIFFMM, "Detailed findings", para. 1432.

The connection between anti-Rohingya content online and the enactment of offline violence is an important part of any assessment of Meta's responsibility for human rights abuses suffered by the Rohingya. As has been demonstrated, the mass dissemination of messages that advocated hatred inciting violence and discrimination against the Rohingya, as well as other dehumanizing and discriminatory anti-Rohingya content, was instrumental in increasing the likelihood of mass violence against the target group, with severe consequences for a wide range of human rights. This assessment is explored in detail in Chapter 8, below.

5.4 CONTENT MODERATION AND ANTI-ROHINGYA CONTENT ON FACEBOOK

"Content moderation" refers to social media platforms' oversight and enforcement of platform rules in relation to permissible and prohibited forms of expression. For Meta, these rules are known as "Community Standards". This section provides an overview and analysis of Meta's efforts to enact effective content moderation in the context of rising levels of anti-Rohingya content on the Facebook platform before and during the 2017 atrocities in Northern Rakhine State.

Content moderation typically involves the detection of harmful content (such as "hate speech") and then a range of varied actions in response, such as removal of the content, making the content less visible ('demoting' or 'downranking'), or penalties against individual users or groups. Content moderation is enacted via a combination of human content moderators and content moderation algorithms.[161]

Meta has long had internal policies against the use of "hate speech" on its platform. The company's community standards recognize the fact that hate speech on the platform can lead to offline violence and harms, stating: "…[W]e don't allow hate speech on Facebook. It creates an environment of intimidation and exclusion, and in some cases may promote offline violence".[162] In 2018, as Meta faced increasing scrutiny for its role in the commission of crimes under international law against the Rohingya, Mark Zuckerberg was questioned in a hearing at the US Senate about the company's role in the possible genocide against the Rohingya. He responded:

> **"Senator, what's happening in Myanmar is a terrible tragedy, and we need to do more … we're hiring dozens of more Burmese-language content reviewers, because hate speech is very language-specific. It's hard to do it without people who speak the local language, and we need to ramp up our effort there dramatically."[163]**

Meta's failure to invest in adequate content moderation in Myanmar has been heavily criticized as new details have come to light. According to one investigation, Meta had only five Burmese speakers to monitor and moderate content for Myanmar's 18 million Facebook users at the time of Zuckerberg's testimony to the US Senate in April 2018, and none of the native speakers were based in Myanmar.[164] Even this small number of staff marked a significant improvement on the situation as it stood prior to the 2017 crisis. In mid-2014, Meta staff admitted in a secret Facebook group containing a number of Myanmar civil society activists that they only had one single Burmese-speaking content moderator devoted to Myanmar at the time, based in their Dublin office.[165]

Research conducted by Amnesty International provides further evidence of the inadequacy of Meta's staffing of its Myanmar content team during the Rakhine crisis. Amnesty International interviewed six Rohingya refugees who recalled how they tried to 'report' anti-Rohingya content on Facebook before and during the 2017 atrocities, only for their reports to either be ignored or to be told that the content did not violate

[161] For a detailed overview of content moderation from a human rights perspective, see: Ranking Digital Rights, 'It's Not Just the Content, It's the Business Model: Democracy's Online Speech Challenge', 17 Mach 2020, newamerica.org/oti/reports/its-not-just-content-its-business-model

[162] Facebook's definition of "hate speech" differs from the definition of "advocacy of hatred" under Article 20 of the ICCPR, with Facebook's definition encompassing a broader range of expression that would not always reach the threshold in Article 20. In June 2017, Facebook stated that its definition of hate speech at the time was "anything that directly attacks people based on what are known as their 'protected characteristics'" and affirmed that it did not allow hate speech on the platform. See Meta, *Hard Questions: Who Should Decide What Is Hate Speech in an Online Global Community?,* 27 June 2017, about.fb.com/news/2017/06/hard-questions-hate-speech

[163] The Washington Post, "Transcript of Mark Zuckerberg's Senate Hearing", 10 April 2018, washingtonpost.com/news/the-switch/wp/2018/04/10/transcript-of-mark-zuckerbergs-senate-hearing

[164] Cecilia Kang and Sheera Frenkel, *An Ugly Truth: Inside Facebook's Battle for Domination*, Chapter 9, p.191.

[165] Referenced in Cecilia Kang and Sheera Frenkel, *An Ugly Truth: Inside Facebook's Battle for Domination*, and confirmed by an Amnesty International interview with one activist who was in the Facebook group.

Facebook's community standards.[166]

Sharif, a 28-year-old Rohingya community educator, told Amnesty International that he reported anti-Rohingya content "more than 100 times" since 2014, and no action was ever taken.[167] Showkutara had a similar experience, and recalled her dismay at Meta's repeated failure to take action in response to her reports:

> **"There were so many pages and contents, how could I report them all? I was not able to do anything against all these things. But I did report some, and I just received a message [that no action would be taken]… I really wanted to stop these things on Facebook, and I tried a lot – I just cried when I saw this, I didn't have any other option. I really wanted Facebook to stop this hate speech spreading, but I could not, and it made me so upset."[168]**

Civil society activists working on digital rights in Myanmar at the time of the 2017 atrocities in northern Rakhine State offered a similar account. Victoire Rio, who headed social impact for the Phandeeyar foundation – a technology and innovation hub – at the time, recounted discussing the inadequacies of Meta's content moderation in a meeting with a delegation of Meta staff to Myanmar in June 2017:

> "I raised the fact that we needed a much better escalation mechanism to address emergencies, and that the reporting system was not working. I was told that the on-platform report function was the fastest way to get something reviewed, even in emergencies. I was told this took six hours on average – and I laughed out loud at the guy, as my experience was that the reports took far longer to receive a response, if they ever got one at all, and the response received was typically 'this is not a violation of community standards'.
>
> We actually ran an experiment to test the turnaround time at scale following that meeting, and found that the large majority of the content, if it ever was reviewed, took over 48 hours to get a response. In many cases, the response came within minutes of 48 hours, suggesting that was likely Facebook's internal target. But a post could go very viral within 48 hours."[169]

In 2018 group of civil society activists wrote an open letter to Mark Zuckerberg about viral chain messages which had been circulated via Facebook Messenger and which sought to stoke violence between Buddhists and Muslims. The letter described these messages as "clear examples of your tools being used to incite real harm" and complained that "far from being stopped, they spread in an unprecedented way, reaching country-wide and causing widespread fear and at least three violent incidents in the process."[170]

Even after Meta's role in respect of the violence in Rakhine state had received global attention, Meta's response remained wholly inadequate. An August 2018 analysis of over 1,000 Facebook posts by Reuters and the Human Rights Center at UC Berkeley revealed that, one year after the peak of the violence against the Rohingya in Rakhine State, Meta had done little to address the huge amount of harmful anti-Rohingya content on the platform.[171]

Meta's failures in respect of content moderation were also the focus of significant criticism from the IIFFMM. In its final report, the IIFFMM (or "the Mission") noted, "[a]lthough improved in recent months, the response of Facebook has been slow and ineffective".[172] The Mission detailed its own struggles to seek an appropriate response from Meta after witnesses it engaged with faced death threats on Facebook due to their cooperation with the Mission. The relevant extract is contained in full below:

> "The Mission itself experienced a slow and ineffective response from Facebook when it used the standard reporting mechanism to alert the company to a post targeting a human rights defender for

[166] Amnesty International interview by video call with Maung Sawyeddollah, 6 April 2022; Amnesty International interview by video call with Tun (pseudonym), 11 April 2022; Amnesty International interview by video call with Mohammed Junaid, 18 April 2022; Amnesty International video call with Mohammed Showife, 12 April 2022; Amnesty International interview by video call with Showkutara, 6 April 2022.

[167] Amnesty International Interview with Mohamed Sharif (pseudonym), 13 April 2022.

[168] Amnesty International interview by video call with Showkutara, 06 April 2022.

[169] Amnesty International interview by video call with Victoire Rio, 12 April 2022.

[170] Civil Society Open Letter to Facebook, Yangon Myanmar, 5 April 2018, https://drive.google.com/file/d/1Rs02G96Y9w5dpX0Vf1LjWp6B9mp32VY-/view

[171] Reuters, "Why Facebook is losing the war on hate speech in Myanmar", 15 August 2018, https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/

[172] IIFFMM, "Final Report", para 74.

> his alleged cooperation with the Mission. The post described the individual as a "national traitor", consistently adding the adjective "Muslim". It was shared and re-posted over 1,000 times.
>
> Numerous comments to the post explicitly called for the person to be killed, in unequivocal terms: "Beggar-dog species. As long as we are feeling sorry for them, our country is not at peace. These dogs need to be completely removed." "If this animal is still around, find him and kill him. There needs to be government officials in NGOs." "Wherever they are, Muslim animals don't know to be faithful to the country." "He is a Muslim. Muslims are dogs and need to be shot." "Don't leave him alive. Remove his whole race. Time is ticking."
>
> The Mission reported this post to Facebook on four occasions; in each instance the response received was that the post was examined but "doesn't go against one of [Facebook's] specific Community Standards". The Mission subsequently sent a message to an official Facebook email account about the matter but did not receive a response. The post was finally removed several weeks later but only through the support of a contact at Facebook, not through the official channel.
>
> Several months later, however, the Mission found at least 16 re-posts of the original post still circulating on Facebook. In the weeks and months after the post went online, the human rights defender received multiple death threats from Facebook users, warnings from neighbours, friends, taxi drivers and other contacts that they had seen his photo and the posts on Facebook, and strong suggestions that the post was an early warning. His family members were also threatened. The Mission has seen many similar cases where individuals, usually human rights defenders or journalists, become the target of an online hate campaign that incites or threatens violence.'[173]

In several ways, the IIFFMM's experience is emblematic of Meta's failures in respect of content moderation in Myanmar before, during, and even after the 2017 atrocities. The fact that the same UN mission which was known to be investigating Meta's role in a possible genocide could not succeed in their repeated efforts to have content removed – even though the content clearly incited violence and discrimination against the Rohingya – speaks volumes about the extent of the company's failings.

As the Mission sought in vain to report the offending post and have it taken down, the post had already gone viral and rapidly spread on the Facebook platform. This dynamic speaks to the inherent limitations of content moderation – even when it is better resourced – as a solution to the rapid spread of harmful content at scale on social media platforms, which is often boosted by the content-shaping algorithms underpinning the platforms themselves. This limitation is analysed in detail in Section 6.1, below.

Sawyeddollah, a 21-year-old Rohingya refugee and youth activist, recalled how his frustrations at Meta's content moderation failures led him to believe that Meta itself was also contributing to the suffering of his community:

> **"I saw a lot of horrible things on Facebook. And I just thought that the people who posted that were bad. I didn't think that Facebook was to blame. But one day I saw a post that made me feel so bad.**
>
> **It said, 'these Bengali people' - using derogatory words - 'their birth rate is so much higher than us – if they live on, we will be under their rule soon'. I tried to report that post to Facebook. I said it was hate speech. But I got a response that said thank you for reporting it, but sorry, it does not go against community standards. That made me really angry. Then I realized that it is not only these people – the posters – but Facebook is also responsible. Facebook is helping them by not taking care of their platform."[174]**

After the 2014 Mandalay riots triggered the Myanmar authorities to temporarily block Facebook in 2014, the company moved to support a civil society-led, anti-hate-speech initiative known as 'Panzagar' or 'flower speech' by creating a 'flower speech sticker pack'.[175] The idea of the initiative was to enable users to post 'counter-speech' whenever they saw content that advocated violence or discrimination on the platform. The stickers contained messages such as "Think before you share" and "Don't be the cause of violence."[176] However, this initiative may have had some serious unintended consequences, as described in Section 6.3, below.

Myanmar digital rights activists interviewed by Amnesty International shared their view that Meta has

[173] IIFFMM, "Detailed findings", para. 1351.

[174] Amnesty International interview by video call with Sawyeddollah, 06 April 2022.

[175] See: "Support Panzagar: End Hate Speech with Flower Speech", Facebook Page, facebook.com/supportflowerspeech

[176] Sheera Frenkel & Cecila Kang, *An Ugly Truth: Inside Facebook's Battle for Domination,* 2021, p. 192.

improved its civil society engagement and some aspects of its content moderation practices in Myanmar in more recent years.[177] The IIFFMM also recognized improvements since April 2018, noting Meta's ban of Ma Ba Tha and a number of individual military officers, such as Min Aung Hlaing, and civilian figures spreading hate, including Wirathu (the radical anti-Rohingya monk described in Section 5.1, above).[178] Following the coup in January 2021, Meta announced that it would ban the Myanmar military from the platform.[179]

These improvements in Myanmar have not occurred across all of Meta's products and services; however. Meta has increasingly turned to AI systems to identify and moderate harmful content, including "hate speech", on its platforms. An investigation by Global Witness in March 2022 found that Meta's content moderation algorithms were still failing to detect blatant anti-Rohingya and anti-Muslim content on the platform. The organization collated eight examples of "hate speech" directed against the Rohingya from the IIFFMM's reports and submitted each of these examples to Meta in the form of advertisements in Burmese. All eight of the adverts were accepted by Meta for publication on Facebook.[180] This study underscored the company's continued challenges in respect of content moderation, even in a country where it has invested so much.

The Facebook Papers – a cache of internal Meta documents which were disclosed by whistle-blower Frances Haugen to the US Congress in October 2021 – have also shown that Meta has failed to adequately invest in content moderation in many other parts of the Global South.[181] The company heavily prioritized English-speaking markets, especially the United States, and often neglected its responsibilities in less politically influential countries. Its AI-driven content moderation classifiers have been shown to be woefully inadequate in languages other than English.[182] Although incidents such as the Cambridge Analytica scandal and the 6 January Capitol Hill riots have dominated the narrative about Meta's failings – in reality, the company's human rights failures have disproportionately impacted countries in the Global South.

One former Meta employee told Amnesty International that the approach internally at Meta mirrored the relative indifference of Western media to lives in the Global South: "Different countries are treated differently. If 1,000 people died in Myanmar tomorrow, it is less important than if 10 people in Britain die". In relation to the matter of staffing for content moderation, the former employee added, "I think Facebook is really reluctant to hire specific staff for every one of the 195 countries in the world – so they only staff 'important' countries – which runs into serious problems."[183]

As outlined in this section, Meta consistently failed to enforce its content policies to appropriately address widespread anti-Rohingya content on the Facebook platform in the months and years preceding the 2017 atrocities in northern Rakhine State. Meta's failure to enforce its own community standards is an important consideration in the assessment of the adequacy of Meta's human rights due diligence measures in Myanmar, as discussed further in Chapter 8, below.

[177] Amnesty International interview by video call with Victoire Rio, 12 April 2022; Amnesty International interview by video call with Zaw (pseudonym), 28 March 2022.

[178] IIFFMM, "Detailed findings", para 1353.

[179] Billy Perrigo, 'Facebook's Ban of Myanmar's Military Will Be a Test of the True Power of Social Media Platforms', 1 March 2021, Time, time.com/5943151/facebook-myanmar-military-ban

[180] Global Witness, 'Facebook approves adverts containing hate speech inciting violence and genocide against the Rohingya', 20 March 2022, globalwitness.org/en/campaigns/digital-threats/rohingya-facebook-hate-speech

[181] See: Rest of the World, "The Facebook Papers reveal staggering failures in the Global South", 26 October 2021, restofworld.org/2021/facebook-papers-reveal-staggering-failures-in-global-south

[182] AP News, "Facebook's language gap weakens screening of hate, terrorism", 25 October 2021, apnews.com/article/the-facebook-papers-language-moderation-problems-392cb2d065f81980713f37384d07e61f

[183] Amnesty International interview by video call with a former Meta employee, 27 April 2022.

6. “THE MECHANICS OF OUR PLATFORM ARE NOT NEUTRAL”: HOW FACEBOOK FUELED HUMAN RIGHTS HARMS AGAINST THE ROHINGYA

> **“We have evidence from a variety of sources that hate speech, divisive political speech, and misinformation on Facebook and the family of apps are affecting societies around the world. We also have compelling evidence that our core product mechanics, such as virality, recommendations, and optimizing for engagement, are a significant part of why these types of speech flourish on the platform… The mechanics of our platform are not neutral.”**
>
> Unnamed Facebook employee, 12 August 2019[184]

Amnesty International and others have previously warned of the human rights risks associated with Meta’s use of content-shaping algorithms (as defined in Section 4.2, above).[185] These algorithms are based on the premise that the more engaged an individual is on Facebook, the longer they will spend on the platform, which increases their exposure to tracking-based advertising and boosts Meta’s revenues. Equally, the more time users spend on Facebook, the more invasive personal data the company can collect on them. This data enables enhanced micro-targeting to increase revenues. This model has driven Meta to constantly seek out new sources of personal data and new markets across the globe. The model has proven to be extraordinarily profitable for Meta.

When whistle-blower Frances Haugen testified before the U.S. Congress in October 2021 on the most harmful aspects of Meta’s business practices, much of her testimony focused on Meta’s use of content-

[184] The Facebook Papers, “What is Collateral Damage?”,12 August 2019, comment on p. 34.

[185] Amnesty International, *Surveillance Giants: How the business model of Google and Facebook threatens human rights”* (Index: POL 30/1404/2019), 21 November 2019, amnesty.org/en/documents/pol30/1404/2019/en, p. 34.

shaping algorithms to proactively promote certain types of harmful content in the pursuit of profit.[186] Her testimony was informed by a huge cache of internal Meta documents which Haugen shared with members of the U.S. Congress. Amnesty International has reviewed many of these documents in detail – known as "the Facebook Papers" – and conducted an analysis of their human rights implications in respect of Meta's duty to provide a remedy to the Rohingya. This analysis – informed by interviews with subject matter experts and other sources – is contained throughout this chapter.

6.1 "PLAYING WHACK-A-MOLE": THE FALLACY OF CONTENT-BASED SOLUTIONS TO ALGORITHMIC HARMS

> **"We are never going to remove everything harmful from a communications medium used by so many, but we can at least do the best we can to stop magnifying harmful content by giving it unnatural distribution."**
>
> Extract from an internal Meta memo, December 2019[187]

As demonstrated in Section 5.4, above, Meta consistently failed to enforce its own content policies in respect of "hate speech" against the Rohingya in Myanmar. However, even if the company had conducted more effective content moderation, these measures may not have sufficed to mitigate the human rights risks and harms caused by the company's business practices. This section assesses the limitations of content-based solutions (such as content moderation) to harms resulting from content-shaping algorithms.

In 2018, alongside the publication of its human rights impact assessment for Myanmar, Meta partially acknowledged its role in the mass violence perpetrated against the Rohingya. The company stated: "prior to this year, we weren't doing enough to help prevent our platform from being used to foment division and incite offline violence. We agree that we can and should do more."[188]

This statement reflects Meta's position that its primary failing was its inadequate moderation of content posted by other actors on the platform. This position effectively covers up Meta's proactive role in the amplification of anti-Rohingya content, and it seeks to cast content moderation as the main solution to the spread of harmful content on the Facebook platform. But there are several problems with this position, most notably the fact that Meta's own content-shaping algorithms promote and fuel the spread of harmful content, including content that advocates hatred (as detailed in Section 6.2, below). These content-shaping algorithms – which are optimized to maximize user engagement and therefore profit – are central to Meta's overall business model.[189] As a result, content moderation fails to address the root cause of Meta's "hate speech" problem.

Michael, the international aid worker who was subject of a viral Facebook post including threats to his life, noted the near impossibility of taking action against every single post among a tide of viral content:

> **"My friends in Myanmar and in the US were flagging and reporting these posts. Many of them were being taken down, as far as I know. But the one with my passport photo had been shared over 1000 times by then… My friends had been reporting hundreds of different posts, it was like a game of whack-a-mole."[190]**

Recognizing these dynamics, Meta whistle-blower Frances Haugen stated during her testimony to the U.S.

[186] See: Karen Hao, "The Facebook whistleblower says its algorithms are dangerous. Here's why.", *MIT Technology Review*, 5 October 2021, technologyreview.com/2021/10/05/1036519/facebook-whistleblower-frances-haugen-algorithms

[187] The Facebook Papers, "We are Responsible for Viral Content", 11 December 2019, p.17, emphasis added.

[188]Meta, "An Independent Assessment of the Human Rights Impact of Facebook in Myanmar", 5 November 2018, about.fb.com/news/2018/11/myanmar-hria

[189] For a thorough analysis of the inadequacy of content moderation as a solution for business model-related harms, including the algorithmic amplification of problematic content, see: Ranking Digital Rights, 'It's Not Just the Content, It's the Business Model: Democracy's Online Speech Challenge', 17 Mach 2020, available at newamerica.org/oti/reports/its-not-just-content-its-business-model; and Ranking Digital Rights, Getting to the Source of Infodemics: It's the Business Model', 27 May 2020, available at newamerica.org/oti/reports/getting-to-the-source-of-infodemics-its-the-business-model

[190] Amnesty International interview by video call with Michael (pseudonym), 24 March, 2022.

Congress: "I'm a strong advocate for non-content-based solutions, because those solutions will protect the most vulnerable people in the world," making reference to the limits of content moderation as a solution.[191] Non-content-based solutions include changes to Meta's content-shaping algorithms. Haugen has specifically suggested that Meta should "get rid of the engagement-based ranking" and advocated for a return to Meta's chronological news feed.[192]

The size and scale of Meta, and its presence in almost every country in the world, add to the difficulties associated with effective content moderation. These structural limitations have been compounded by the fact that Meta has, as previously noted, failed to adequately invest in resourcing its operations in the non-English speaking world, enabling algorithmically amplified "hate speech" to run riot in the absence of effective moderation.

In October 2020, Meta CEO Mark Zuckerberg testified to the US Senate that:

> **"[O]ver the last few years, we've gone from proactively identifying and taking down about 20% of the hate speech on the service, to now, *we are proactively identifying, I think it's about 94% of the hate speech that we ended up taking down* [emphasis added], and the vast majority of that before people even have to report it to us."[193]**

However, internal company records contained in the Facebook Papers reveal this to be a serious mischaracterisation of the reality of Meta's success at effectively moderating "hate speech" on the platform. An internal document entitled 'Demoting on Integrity Signals is not enough', which appears to have been shared internally in July 2019,[194] contains the revelation that: "…according to best estimates I have seen, we only take action against approximately 2% of the hate speech on the platform".[195]

These 94% and 2% figures are not necessarily contradictory as they refer to different metrics. The 94% figure refers to the amount of "hate speech" proactively identified by Meta's content moderation algorithms, where the remainder (6%) refers to the amount of "hate speech" the company identifies through other means (such as reporting by users). This metric does not measure or account for "hate speech" *not* identified by Meta. The 2% figure, on the other hand, refers to the proportion of hate speech actually existing on the Facebook platform which is actioned by Meta – with the remaining 98% representing the total share of "hate speech" which is not addressed by the company.

Meta has moved to publicly defend its record in relation to "hate speech" by claiming that the metrics it uses are most relevant when measuring this form of harmful content on their platform.[196] However, as some critics have pointed out, "what matters to society is the amount of hate speech that is *not* removed from the platform" – and not the relative improvement in Meta's content moderation algorithms at detecting "hate speech", which Meta's preferred metric refers to.[197] Meta's efforts to focus on alternative metrics in relation to hate speech has also been criticized as being "emblematic of a long-standing strategy by Facebook to obfuscate and omit information in transparency reports and other public statements".[198]

The same document from the Facebook Papers further recognizes the limitations of Meta's content moderation algorithms at capturing violating content such as hate speech, noting:

[191] Karen Hao, "The Facebook whistleblower says its algorithms are dangerous. Here's why.", *MIT Technology Review*, 5 October 2021, technologyreview.com/2021/10/05/1036519/facebook-whistleblower-frances-haugen-algorithms

[192] Karen Hao, "The Facebook whistleblower says its algorithms are dangerous. Here's why.", *MIT Technology Review*, 5 October 2021, technologyreview.com/2021/10/05/1036519/facebook-whistleblower-frances-haugen-algorithms

[193] Rev, "Tech CEOs Senate Testimony Transcript October 28", 28 October 2020, rev.com/blog/transcripts/tech-ceos-senate-testimony-transcript-october-28, emphasis added.

[194] The document is undated, but the comments on the document are all from July 2019. The Facebook Papers, "Demoting on Integrity Signals is not enough", undated.

[195] The Facebook Papers, "Demoting on Integrity Signals is not enough", undated, emphasis added.

[196] Guy Rosen, "Hate Speech Prevalence Has Dropped by Almost 50% on Facebook", *Meta*, 17 October 2021, fb.com/news/2021/10/hate-speech-prevalence-dropped-facebook ; Ranking Digital Rights, "Cross-checking Facebook: Five Lies Revealed by Frances Haugen", 14 October 2021, rankingdigitalrights.org/2021/10/14/cross-checking-facebook-frances-haugen

[197] Noah Giansiracusa, "Facebook Uses Deceptive Math" (previously cited), 15 October 2021.

[198] Ranking Digital Rights, "Cross-checking Facebook" (previously cited).

> **“the problem is that we do not and possibly never will have a model that captures even a majority of integrity harms [i.e., the spread of harmful content on the platform], particularly in sensitive areas”.[199]**

Underlining the extent to which these limitations in addressing “hate speech” are a result of Facebook’s core features, including news feed ranking, the author of the document states:

> **“This leaves us in a difficult position. We do have reasonable metrics that can tell us when a given ranking change is likely to cause integrity harms – even with low precision and recall, we can get a decent sense of whether a launch is increasing hate speech, or misinformation, or other harms. However, we don’t have a way of effectively demoting this content in a targeted way (i.e. without significant collateral damage); and even if we did, we often won’t be able to launch them based on policy concerns”.[200]**

Here, the author of this document suggests that Meta staff have the means to predict when tweaks to their content-shaping algorithms (such as those aimed at maximizing engagement) are likely to result in the amplification of “hate speech” and other harmful content. The author also admits, however, that Meta does not have the capacity to effectively reduce the amplification of harmful content in a narrow and targeted manner. They add that in any case, “policy concerns” often rule out any technical interventions which are aimed at reducing the spread of harmful content. It is not clear precisely which “policy concerns” are referred to in this document; however, this term in the Facebook Papers often refers to political considerations which influence company policy, such as whether a specific government or political actor will perceive an action by the company as biased.

This logic of content-based solutions to algorithmically amplified harms rests on the premise that Meta is a neutral arbiter of content. In reality, Meta’s business model, along with its core products and services, fuel the spread of harmful content. Meta’s surveillance-based business model incentivizes the spread of maximally engaging, inflammatory content, including advocacy of hatred. These content-shaping algorithms which underpin Meta’s business model are analysed in further detail in Section 6.2, below.

6.2 ENGAGEMENT AT ALL COSTS: HOW META MAKES HATE GO VIRAL

> **“We [Facebook] have evidence from a variety of sources that hate speech, divisive political speech, and misinformation on Facebook and the family of apps are affecting societies around the world. We also have compelling evidence that our core product mechanics, such as virality, recommendations, and optimizing for engagement, are a significant part of why these types of speech flourish on the platform”**
>
> A Meta employee commenting in a leaked internal company document, August 2019 [201]

The Facebook Papers reveal in greater detail than ever before how human rights risks and harms are driven by Meta’s business model and its core content-shaping algorithms. This section provides an in-depth analysis of how these algorithms operate in practice, and how they fuel human rights harms.

6.2.1 VIRALITY

One of Facebook’s most well-known features is algorithmic virality – whereby certain content is algorithmically boosted in such a way that enables it to reach a far wider audience than it could otherwise reach ‘organically’ (without being boosted). Another document contained in the Facebook Papers with potentially significant human rights implications is a post, entitled, “We Are Responsible for Viral Content”.

[199] The Facebook Papers, “Demoting on Integrity Signals is not enough”, undated.

[200] The Facebook Papers, “Demoting on Integrity Signals is not enough”, undated.

[201] The Facebook Papers, “What is Collateral Damage?”, 12 August 2019, p. 34.

The post is written by an unknown employee and dated 11 December 2019. It explains Meta's responsibility for viral content in clearly understandable terms:

> **"Unlike communication with close friends and family, virality is something new we have introduced to many ecosystems… and *it occurs because we intentionally encourage it for business reasons*"[202]**

The post acknowledges that:

> **"[R]anking content about higher stakes topics like health or politics based on engagement leads to perverse incentives and integrity issues."[203]**

In a clear articulation of how Meta's algorithmic ranking systems operate in practice, combined with a recognition of Meta's responsibility for the harms it produces, the author states:

> **"Our ranking systems have specific separate predictions for not just what you would engage with, but what we think you may pass along so that others may engage with. Unfortunately, research has shown how *outrage and misinformation are more likely to be viral,* and recent experiments that deprecate these models indicate that removing these models does positively impact metrics for misinformation and hate".[204]**

The post further reveals that the viral content shown to Facebook users is not solely based on users' preferences, but rather targeted in a manner which would maximize engagement by stoking emotive responses. It also gives an insight into how 'engagement' is measured and rewarded on Facebook:

> **"We prioritize content based on engagement, but engagement doesn't necessarily mean that a user actually wants to see more of something. One of our biggest signals we use to provide more similar content is comments and… a comment that you hate a thing can be seen as a positive signal leading content to get outsized distribution. People game this in various ways, posting ever more outrageous things to get comments and reactions that our algorithms interpret as signs we should allow things to go viral*…*".[205]**

The post clearly acknowledges how Meta's optimization of its platform for maximizing engagement in the pursuit of maximum profit carries significant systemic risks in the societies in which Meta operates: "…[A]s long as we continue to optimize for overall engagement and not solely what we believe individual users will value, we have an obligation to consider what the effect of optimizing for business outcomes has on the societies we engage in." [206]

With specific reference to high-risk environments, such as Myanmar, the author highlights the risks which this model poses to ethnic groups at risk of discrimination or violence: "Recent research has shown how misinformation and divisive messages travel faster on social media… What happens when you introduce virality into places where people will believe anything or in places where people are pre-disposed to believe the worst about vulnerable ethnic groups? The honest answer is we don't know."[207]

The author of the document offers some suggestions to address the risks and harms associated with engagement-centric virality:

> **"Since we have introduced virality to the world, our integrity efforts should focus on making virality a force for good – or at least less bad. We are never going to remove everything harmful from a communications medium used by so many, *but we can at least do the best we can to stop magnifying harmful content by giving it unnatural distribution* [emphasis added]. We should also take seriously the charge that we are affecting media ecosystems by creating perverse incentives."[208]**

Although these warnings and recommendations were authored in 2019, Meta appears to have done little to change the fundamental design and orientation of its content-shaping algorithms since then. Meta's failure to act on warnings such as these led whistle-blower Frances Haugen to testify before the U.S. Senate in

[202] The Facebook Papers, "We are Responsible for Viral Content", 11 December 2019, emphasis added.

[203] The Facebook Papers, "We are Responsible for Viral Content", 11 December 2019, p.26.

[204] The Facebook Papers, "We are Responsible for Viral Content", 11 December 2019, emphasis added.

[205] The Facebook Papers, "We are Responsible for Viral Content", 11 December 2019, p.8.

[206] The Facebook Papers, "We are Responsible for Viral Content", 11 December 2019, p.7.

[207] The Facebook Papers, "We are Responsible for Viral Content", 11 December 2019, p.17.

[208] The Facebook Papers, "We are Responsible for Viral Content", 11 December 2019, p.17, emphasis added.

October 20201 that “Facebook’s products harm children, stoke division, and weaken our democracy”.[209] Amnesty International wrote to Meta in May 2022 to ask when the company first became aware of the risks posed by its ranking and recommendation algorithms, and what actions Meta has taken to address such risks. Meta did not answer this question in its response.

6.2.2 NEWS FEED & RANKING

Meta’s News Feed algorithm is central to what content is seen by whom on the Facebook platform. It curates a highly personalized user experience based on invasive profiling enabled by the collection of intimate personal data. It ranks specific types of content – deciding what users see first when they open Facebook. An internal company document dated 9 March 2020 explains how individual users see specific pieces of content based on ranking scores assigned by Meta’s ranking algorithms:

> **“News feed ranking is another way Facebook becomes actively involved in these harmful experiences. Of course users also play an active role in determining what content they are connected to through feed, by choosing who to friends and follow. Still, when and whether a user sees a piece of content is also partly determined by the ranking scores our algorithms assign, which are ultimately under our control.”[210]**

The same document reveals that Meta staff considered the ethical implications of the company’s use of content-shaping algorithms. The document, entitled, ‘Facebook and Responsibility’, stands in contrast to Meta’s public position on its responsibility for promoting content that resulted in offline violence. In a stunning recognition of responsibility, the document states:

> **“Actively ranking content in News Feed and promoting content on recommendations surfaces makes us *responsible for any harm caused by exposure to that content*” [emphasis added].[211]**
>
> **“This means, according to ethicists, Facebook is always at least *partially responsible for any harmful experiences on News Feed*. This doesn’t owe any flaw with our News Feed ranking system, it’s just inherent to the process of ranking …. Crucially, all of these algorithms produce a single score; a score Facebook assigns. Thus, there is no such thing as inaction on Feed, we can only *choose to take different kinds of actions”* [emphasis added].[212]**

6.2.3 RECOMMENDATIONS FOR USER BEHAVIOUR AND ACTIONS

Meta also deploys its algorithmic technologies to actively promote and recommend certain actions and behaviours to Facebook users. Recommendations are algorithmically delivered to users on their feeds, and include ‘groups you should join’, ‘people you may know, and ‘pages you should follow’. The Facebook Papers provide additional detail in relation to how Meta recommends content on Facebook:

> **“Facebook is most active on delivering content to users on recommendation surfaces like ‘Pages you may like’, ‘Groups you should join’ and suggested videos on Watch… Users don’t opt in to these experiences by following other users or pages. Instead, *Facebook is actively presenting these experiences, and according to ethicists, is therefore entirely responsible if these experiences are harmful* [emphasis added].”[213]**

[209] Karen Hao, “The Facebook whistleblower says its algorithms are dangerous. Here’s why.”, *MIT Technology Review*, 5 October 2021, technologyreview.com/2021/10/05/1036519/facebook-whistleblower-frances-haugen-algorithms

[210] The Facebook Papers, “Facebook and Responsibility”, 09 March 2020, documentcloud.org/documents/21594152-tier2_rank_other_0320

[211] The Facebook Papers, “Facebook and Responsibility”, 09 March 2020, documentcloud.org/documents/21594152-tier2_rank_other_0320

[212] The Facebook Papers, “Facebook and Responsibility”, 09 March 2020, pp. 4-5. documentcloud.org/documents/21594152-tier2_rank_other_0320

[213] The Facebook Papers, “Facebook and Responsibility”, 09 March 2020, documentcloud.org/documents/21594152-tier2_rank_other_0320

The author of this internal company post goes on to ponder on a hypothetical scenario reflecting upon Meta's ethical responsibility for actively promoting certain actions and types of content:

> **"Nearly everyone judges that it is better to passively let one person die than to actively kill another person to save them – even though both action and inaction result in one life lost (Borg et al., 2006). But, do users understand the active role that Facebook plays in making recommendations and ranking content?"[214]**

Here, the author of the document recognizes that Meta's use of content-shaping algorithms to actively amplify and promote harmful content bestows a heightened responsibility on the company for any harms caused by that content. This responsibility is far higher than if Meta were only failing to effectively moderate and remove harmful content posted by others. The author of the post goes on to make an impassioned case for a change in Meta's approach to the harms caused by its content-shaping algorithms:

> **"I think appreciating that we are responsible for harmful experiences on any surface where we actively present content makes clear what we cannot do: we cannot stand by without acting or pretend that we are bystanders to what transpires on these surfaces."[215]**

6.2.4 GROUPS

Another feature which has been identified as posing specific risks in respect of advocacy of hatred that incites violence and discrimination is Meta's 'Groups' feature on Facebook. According to Meta, Groups are "a place to communicate about shared interests with certain people", and users can "create a group for anything."[216]

In 2016, internal Meta research reported on by the Wall Street Journal found "extremist content thriving in more than one-third of large German political groups on the platform", and most significantly, it also recognized that Meta's content-shaping algorithms were responsible for the growth of these groups.[217] The report noted:

> **"The 2016 presentation stated that "64% of all extremist group joins are due to our recommendation tools" and that most of the activity came from the platform's "Groups You Should Join" and "Discover" algorithms: "*Our recommendation systems grow the problem* [emphasis added]."[218]**

Despite the warning signs that were emerging from internal company research, Meta doubled down on prioritization of groups in May 2019 as the company unveiled a major pivot towards more 'community', as revealed in a Tech Crunch article:

> **"Everywhere there are friends, there should be Groups," says the head of the Facebook app, Fidji Simo… But Facebook's goal is not only to have 2.38 billion people using the feature – the same number as use its whole app – but to get them all into *meaningful Groups that emblematize their identity* [emphasis added].[219]**

The Facebook Papers enable a deeper understanding than ever before of how Meta's content-shaping algorithms work in practice – and how they promote human rights harms. These damning internal

[214] The Facebook Papers, "Facebook and Responsibility", 09 March 2020, documentcloud.org/documents/21594152-tier2_rank_other_0320, p. 6.

[215] The Facebook Papers, "Facebook and Responsibility", 09 March 2020, documentcloud.org/documents/21594152-tier2_rank_other_0320, p. 8.

[216] Facebook, Help Centre: 'Groups', facebook.com/help/1629740080681586

[217] The Wall Street Journal, "Facebook Executives Shut Down Efforts to Make the Site Less Divisive", 26 May 2020, wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499

[218] The Wall Street Journal, "Facebook Executives Shut Down Efforts to Make the Site Less Divisive", 26 May 2020, wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499

[219]Tech Crunch, "Facebook pivots to what it wishes it was", 2 May 2019, techcrunch.com/2019/05/01/aspirationbook/?guccounter=1, emphasis added

documents reveal how these algorithmic features – including virality, the News Feed, Recommendations, and Groups – all promote and amplify divisive, harmful, and inflammatory content. Nowhere have the consequences of these content-shaping algorithms been more devastating than in the case of the Rohingya.

6.3 HOW META'S ALGORITHMS FUELED ANTI-ROHINGYA VIOLENCE IN MYANMAR

The revelations contained in the Facebook Papers enable a renewed understanding of the manner in which Meta's content-shaping algorithms fuelled the mass violence perpetrated against the Rohingya. In this section, the role of Meta's core content-shaping algorithms is considered in the context of the 2017 atrocities perpetrated against the Rohingya in Northern Rakhine State.

As has been outlined in Section 5.2, above, viral Facebook posts containing dehumanizing narratives and advocacy of hatred that incited violence and discrimination against the Rohingya were a dominant feature of the persecution of the Rohingya before and during the 2017 crisis. They helped to normalize the dehumanization of an entire ethnic group in the most violent, hateful terms imaginable, and facilitated the rapid spread of anti-Rohingya ethnic hatred throughout the country.

In the lead up to the 2017 atrocities, the virality of these posts made them appear ubiquitous on Burmese-language Facebook, creating a sense that everyone in Myanmar shared these views, helping to build a shared sense of urgency in finding a 'solution' to the "Bengali problem"[220] and ultimately building support for the military's 2017 "clearance operations". Michael, the international aid worker who lived in Myanmar from 2013- 2018, described a typical Facebook News Feed in Myanmar in 2017:

> **"The vitriol against the Rohingya was unbelievable online – the amount of it, the violence of it. It was overwhelming. There was just so much. That spilled over into everyday life…**
>
> **The news feed in general [was significant] – seeing a mountain of hatred and disinformation being levelled [against the Rohingya], as a Burmese person seeing that, I mean, that's all that was on people's news feeds in Myanmar at the time. It reinforced the idea that these people were all terrorists not deserving of rights. This mountain of misinformation definitely contributed [to the outbreak of violence]."** [221]

Meta's use of content-shaping algorithms to maximize engagement were not considered by the IIFFMM. According to Chris Sidoti, an expert member of the IIFFMM, the Mission had no information about how these algorithms functioned at the time.[222] He added that his current view is that these algorithms "increase the direct liability of the platform providers for the content that is distributed, because they play a role in bringing inciting and inflammatory material to the users who are most susceptible to being influenced by it".[223]

Despite not overtly looking into Meta's content-shaping algorithms, the IIFFMM's final report nonetheless noted "the high levels of engagement of the followers with the posts (commenting and sharing)".[224] Some of the most egregious posts advocating violence, discrimination, and genocide against the Rohingya which were documented by the IIFFMM were Facebook posts which went viral, garnering up to tens of thousands of interactions and shares on the platform. [225]

There is evidence to suggest that Meta's content-shaping algorithms were amplifying harmful content, including advocacy of hatred, in Myanmar as early as 2014. As noted in Section 5.4, after the Mandalay riots triggered the Myanmar authorities to temporarily block Facebook in 2014, Meta supported a civil society-led

[220] On 1 September 2017, Senior General Min Aung Hlaing stated in a public meeting with ethnic Rakhine leaders: "The Bengali problem was a long-standing one which has become an unfinished job", Simon Lewis, Zeba Siddiqui, Clare Baldwin, Andrew R.C. Marshall, "Special report - The shock troops who expelled the Rohingya from Myanmar", Reuters, 26 June 2018, https://www.reuters.com/article/uk-myanmar-rohingya-battalions-specialre-idUKKBN1JM1YA

[221] Amnesty International interview by video call with Michael (pseudonym), 28 March 2022.

[222] Amnesty International Interview by video call with Chris Sidoti, 13 April 2022.

[223] Amnesty International Interview by video call with Chris Sidoti, 13 April 2022.

[224] IIFFMM, "Detailed findings", para 1311.

[225] IIFFMM, "Detailed findings",, para. 1326.

initiative known as 'Panzagar' or 'flower speech' by creating a 'flower speech sticker pack'.[226] However, the civil society activists involved in the initiative noticed that the stickers were having serious unintended consequences. According to Matt Schissler, as reported by Cecilia Kang and Sheera Frenkel, "Facebook's algorithms counted them as one more way people were enjoying a post. Instead of diminishing the number of people who saw a piece of hate speech, the stickers had the opposite effect of making the posts more popular".[227]

David Madden was among those heavily involved in repeated efforts to convince Meta to take action to address increasing risks in Myanmar between 2015-2017 (as detailed further in Section 7.1, below). He told Amnesty International that although the activist community did not explicitly refer to the dangers associated with Meta's content-shaping algorithms in those exact terms, they were "absolutely attuned to the basic principles of these issues".[228] According to Madden, he and his colleagues "had highlighted to Facebook how quickly dangerous content went viral. It was pretty clear that there were multiple forces at play… There was a system to game. It was often shocking [to observe] the sheer velocity of the spread of this content." [229]

Madden lamented the ineffectiveness of the measures which were offered up by Meta to address the rising tide of anti-Rohingya content and incitement to violence. He reflected: "Well-meaning people with "Panzagar" [flower speech] stickers couldn't beat the algorithm… Someone could threaten the life of a journalist and it could go viral before anyone at Menlo Park had even woken up." [230]

There is evidence which suggests that the key institutional drivers behind advocacy of hatred against the Rohingya on Facebook proactively and successfully leveraged Meta's content-shaping algorithms as they sought to demonize and dehumanize the Rohingya. A 2018 study by Victoire Rio found that the Myanmar military and radical Buddhist nationalist groups such as Ma Ba Tha engaged in "viral engagement tactics":

> **"Tactics such as the use of clickbait headlines, explicit requests for likes and shares, or the tagging of multiple accounts on posts were regularly used by assets believed to belong to these groups' networks. These strategies appear to have been used to maximize newsfeed distribution and boost content reach, as well as to build and maintain an engaged audience."[231]**

6.3.1 RECOMMENDATIONS & NEWS FEED

The Facebook Papers reveal that on an unknown date in 2020, a team within Meta received an "escalation" of a video by the leading anti-Rohingya hate figure U Wirathu. The video was reported for violating community standards. Following an internal investigation, Meta found that its algorithms had been actively promoting the video by this hate figure. The investigation revealed that over 70% of the video's views had come from "chaining, i.e., we are actively recommending divisive and inciting content".[232] According to one definition offered by Meta, "chaining" is an example of "non follower-based distribution" and refers to video content "which auto-plays after a video is complete and suggests what's "Up Next" to viewers".[233] This example exemplifies Meta's active amplification of harmful content in Myanmar. It also highlights the weaknesses of Meta's efforts at improved content moderation in Myanmar as recently as 2020: Wirathu had been ostensibly banned from the platform since 2018.[234]

The same document reveals that Facebook users in Myanmar spend "far more time on video compared with the rest of the world", and "a very large amount of time in chained video… with 40% of video time spent happening on chained video". [235] It also states that "overall video time spent is 53% from chaining".[236] This means that, as of 2020, approximately half of all video content seen by Facebook users in Myanmar was being auto-played for users by Meta's content-shaping algorithms.

[226] See: "Support Panzagar: End Hate Speech with Flower Speech", *Facebook Page*, facebook.com/supportflowerspeech

[227] Sheera Frenkel & Cecilia Kang, *An Ugly Truth: Inside Facebook's Battle for Domination*, 2021, p. 192.

[228] Interview by video call with David Madden, 09 June 2022.

[229] Interview by video call with David Madden, 09 June 2022.

[230] Interview by video call with David Madden, 09 June 2022.

[231] Victoire Rio, "The Role of Social Media in Fomenting Violence: Myanmar".

[232] The Facebook Papers, "Video Time Spent in Myanmar", 12 November 2020, on file with Amnesty International.

[233] Facebook, "Growing Your Live Community On Facebook", facebook.com/fbgaminghome/creators/growing-your-live-community

[234] IIFFM, "Detailed findings", para. 1353.

[235] The Facebook Papers, "Video Time Spent in Myanmar", 12 November 2020, on file with Amnesty International.

[236] The Facebook Papers, "Video Time Spent in Myanmar", 12 November 2020, on file with Amnesty International.

There is evidence to suggest that Meta's algorithms are continuing to amplify harmful content in Myanmar in violation of its own policies. Just one month after Meta banned the Myanmar military from the Facebook platform in 2021, an investigation by Global Witness found that Meta's page recommendation algorithm was still amplifying pro-military content that violated many of its own policies on violence and misinformation.[237]

Michael, the aid worker who received death threats after becoming the subject of a viral Facebook post, recounted how the most inflammatory content seemed to gain the most traction in 2017:

> **"The fact that the comments with the most reactions got priority in terms of what you saw first was big – if someone posted something hate-filled or inflammatory it would be promoted the most – people saw the vilest content the most. I remember the angry reactions seemed to get the highest engagement. Nobody who was promoting peace or calm was getting seen in the news feed at all."[238]**

Facebook groups were also a major concern for local digital rights activists and others concerned about the risks that online vitriol against the Rohingya could spill over into a mass outbreak of violence. Victoire Rio, who worked at the Phandeeyar Foundation before co-founding the Myanmar Tech Accountability Network in early 2018, told Amnesty International:

> **"Our biggest concern was hyper-local groups and pages – they became very popular as people naturally searched for their town… We saw them being used for spreading localized rumours, including unsubstantiated claims of imminent attacks by Muslims. These claims often referred to local landmarks, making them particularly believable and concerning. I also think that the group feature was being used for targeting certain groups."[239]**

Rio cited one example wherein two separate 'chain' messages were circulated among Muslims and Buddhists in Myanmar. Both urged preparedness for planned violence to be perpetrated by the other group ahead on the 2017 anniversary of the 9/11 terrorist attacks in the United States. Rio described how an interfaith activist had received both messages, leading her to believe that recipients of these chain messages were identified from religion-focused Facebook groups.[240] According to reports, these chain messages triggered mobs to attack Muslim houses, shopfronts and a mosque in the towns of Magway and Taungdwingyi.[241]

The importance of groups in the Myanmar context was echoed by others. For digital rights activist Zaw (pseudonym), "groups were the most powerful things that happened on Facebook". He explained how large numbers of people in Myanmar began joining groups from 2014 onwards, and how this fed hateful content: "you would get an alternative reality of the narratives happening in the political arena".[242]

The independent human rights impact assessment for Myanmar commissioned by Meta has been sharply criticized by some human rights experts. Some of the most substantial criticism it has faced relates to the HRIA's failure to account for the impact of Meta's content-shaping algorithms in promoting advocacy of hatred against the Rohingya, in addition to a failure to adequately assess the specific risks presented by the Myanmar context prior to Meta's deployment of certain technologies in the country.[243]

As has been demonstrated in this section, the Facebook platform's signature features – including virality, groups, news feed, and recommendations – played a prominent role in Myanmar's information landscape in the months and years preceding the 2017 atrocities in Northern Rakhine State. At the time, the platform's content-shaping algorithms were largely shielded from scrutiny by a lack of transparency on the part of Meta, meaning that were not directly addressed by either the IIFFMM or the independent HRIA. However, recent revelations make it clear that these features were actively amplifying, promoting and recommending divisive

[237] Global Witness, 'Algorithm of harm: Facebook amplified Myanmar military propaganda following coup', 23 Jun 2021, globalwitness.org/en/campaigns/digital-threats/algorithm-harm-facebook-amplified-myanmar-military-propaganda-following-coup

[238] Amnesty International interview by video call with Michael (pseudonym), 28 March 2022.

[239] Amnesty International interview by video call with Victoire Rio, 12 April 2022.

[240] Amnesty International interview by video call with Victoire Rio, 12 April 2022.

[241] John Reed, "Hate speech, atrocities and fake news: the crisis in Myanmar", *Financial Times*, 22 February 2018, ft.com/content/2003d54e-169a-11e8-9376-4a6390addb44

[242] Amnesty International interview with Zaw (pseudonym), 28 March 2022.

[243] See, for example: Mark Latonero, and Aaina Agarwal, 'Human Rights Impact Assessments for AI: Learning from Facebook's Failure in Myanmar', *Carr Center For Human Rights Policy Harvard Kennedy School*, 2021.

and inflammatory content in Myanmar, in a context where the risk of an outbreak of mass violence against the Rohingya was growing by the day.

These content-shaping algorithms amplified dehumanizing and discriminatory anti-Rohingya content, with direct impacts on the right to equality and the right to be free from discrimination for the Rohingya. They also fuelled offline harms by amplifying content which incited and facilitated violence against the Rohingya, and by delivering this content to the users who were most likely to act upon this incitement and engage in real-world violence. This had knock-on impacts on a range of rights of the Rohingya, including the rights to life and freedom from torture, the right to education, to adequate housing and an adequate standard of living, and the right to food.

6.4 PERILOUS TRANSACTIONS: FINANCIAL INCENTIVES, ADVERTISING, AND ANTI-ROHINGYA HATE

In addition to Meta's content-shaping algorithms, there are other aspects of Meta's business practices which may have resulted in the proliferation of anti-Rohingya content on the Facebook platform. This section first assesses the role of Meta's 'Instant Article' feature in driving financially motivated anti-Rohingya clickbait. It then briefly considers the role of paid advertising purchased by actors who were responsible for driving anti-Rohingya content, including the Myanmar military and Ma Ba Tha.

6.4.1 INSTANT ARTICLE

As noted in Section 5.2, both the Myanmar military and hate groups such as Ma Ba Tha are widely recognized as being the key originators behind anti-Rohingya content on Facebook. However, there is a third category of actor responsible for content production that has been less widely recognized, and whose significance may have been vastly underestimated.

Between 2015-2018, a major shift in Myanmar's online news environment occurred, largely as a result of Meta's global roll-out of 'Instant Article', a platform feature which enabled publishers and content creators to monetize their content.[244] Under the programme, content publishers could receive monetary payments from Meta if they published content directly on the platform, instead of linked to a third-party website. Because financial incentives under Instant Article were linked to the amount of engagement a piece of published content received, publishers were incentivized to re-publish whatever content received the highest levels of engagement. This had the impact of redirecting ad revenues to Meta and away from other websites, thereby boosting Meta revenues. It also supercharged the rise of clickbait articles, sites, and pages in Myanmar - with potentially disastrous consequences for the Rohingya.[245]

An investigation by MIT Technology Review, published in November 2021, revealed the enormous impact this shift had on Myanmar's media landscape. Their research showed that in 2015, before Instant Article was available in the country, six of the ten websites in Myanmar getting the most engagement on Facebook were from 'legitimate media'. By 2017, however – one year after Instant Article was rolled out in the country – legitimate publishers made up only two of the top ten publishers on Facebook, and by 2018, they accounted for zero.[246] According to the investigation, they had been entirely replaced with 'fake news and clickbait websites'.[247] As noted by the IIFFMM, "[f]or many people, Facebook is the main, if not only, platform for online news and for using the Internet more broadly".[248] As such, this new Meta feature for its Facebook platform can be seen to have had an enormous impact on the overall media landscape in the country.

[244] See: Timothy B. Lee, "Facebook Instant Articles, explained", 13 May 2015, *Vox*, vox.com/2015/5/13/8600753/facebook-instant-articles-explained

[245] Karen Hao, 'How Facebook and Google fund global misinformation', *MIT Technology Review*, 20 November 2021, technologyreview.com/2021/11/20/1039076/facebook-google-disinformation-clickbait

[246] Karen Hao, 'How Facebook and Google fund global misinformation', *MIT Technology Review*, 20 November 2021, technologyreview.com/2021/11/20/1039076/facebook-google-disinformation-clickbait

[247] Karen Hao, 'How Facebook and Google fund global misinformation', *MIT Technology Review*, 20 November 2021, technologyreview.com/2021/11/20/1039076/facebook-google-disinformation-clickbait

[248] IIFFMM, "Detailed findings", para. 1345.

The significance of Instant Article for the Rohingya lies in its connection to Meta's content-shaping algorithms. Because financial incentives under Instant Article were linked to the amount of engagement a piece of published content received, publishers were incentivized to re-publish whatever content received the highest levels of engagement. Victoire Rio told Amnesty International how, in the lead up to the 2017 atrocities, "there was a proliferation of clickbait websites – many of which were amplifying anti-Rohingya narratives – such as saying that the Rohingya were having an abnormal number of children, Muslims were hiding weapons in mosques, preparing for jihad, etc."[249]

The MIT Technology Review found that, because of the launch of 'Instant Articles', "[c]lickbait actors cropped up in Myanmar overnight. With the right recipe for producing engaging and evocative content, they could generate thousands of US dollars a month in ad revenue, or 10 times the average monthly salary – paid to them directly by Facebook".[250]

According to Rio, the multiplication of clickbait and sensational content, which was driven by the roll-out of Meta's monetization programme, played an important role in shaping public opinion in favour of the "clearance operations" that took place: "After the [ARSA] attacks, you had the whole machine going into a complete frenzy". Rio told Amnesty International that her research revealed that these clickbait sites were not necessarily run by politically motivated actors: "Some are business-minded folks – including some pro-NLD", underlining the Meta-funded financial motivation behind this development.[251]

The IIFFMM's final report also noted the role of dubious "self-proclaimed" news pages, with huge numbers of followers and high levels of engagement on Facebook, in spreading anti-Rohingya content that incited violence and discrimination.[252]

6.4.2 PAID ADVERTISING

Before and during the 2017 atrocities in Northern Rakhine State, the Facebook platform was also hosting paid advertising from both the Myanmar military and Ma Ba Tha. A study by Victoire Rio found that 'both the military and Ma Ba Tha have been found to make use of Facebook's boosting and ad features'.[253] The examples in Rio's study include paid boosting of the page of Senior General Min Aung Hlaing, who Amnesty International has accused of crimes against humanity and who regularly used his Facebook page to spread dehumanizing and inciting posts targeting the Rohingya.[254] Although Meta banned certain military leaders from the platform in 2018, including Min Aung Hlaing, the company only banned military-linked advertisements in February 2021 following the coup.[255]

Meta's content-shaping algorithms were not the only aspects of the company's business practices to have played a role in the proliferation of hateful and divisive content in Myanmar. As detailed in this section, Meta's drive to expand its advertising revenues also resulted in the creation of financial incentives for the production of increasingly inflammatory content. This occurred against a backdrop of rising anti-Rohingya sentiment and violence in Myanmar, further adding to the echo chamber of anti-Rohingya hatred which was developing on the Facebook platform. At the same time, Meta directly profited from paid advertising by the very actors which were responsible for the demonization, dehumanization – and ultimately the ethnic cleansing – of the Rohingya.

[249] Amnesty International interview by video call with Victoire Rio, 12 April 2022.

[250] Karen Hao, 'How Facebook and Google fund global misinformation', *MIT Technology Review*, 20 November 2021, technologyreview.com/2021/11/20/1039076/facebook-google-disinformation-clickbait

[251] Amnesty International interview by video call with Victoire Rio, 12 April 2022.

[252] IIFFMM, "Detailed findings, para 1312, pp.324-325.

[253] Victoire Rio, 'The Role of Social Media in Fomenting Violence: Myanmar' (previously cited).

[254] See: Section 5.2, above.

[255] Sebastian Strangio, "Facebook Bans Myanmar Military-linked Companies From its Platforms", *The Diplomat*, 9 December 2021, thediplomat.com/2021/12/facebook-bans-myanmar-military-linked-companies-from-its-platforms

7. A DISASTER FOREWARNED: META'S "WILFUL BLINDNESS" IN MYANMAR

> **Why was this not foreseen? How could [Facebook] launch such a massive operation in Myanmar and fail to staff it appropriately? I could never understand this approach. The focus was on extending Facebook and the business. It doesn't align with their human rights [responsibilities] as a corporate actor. It can only be described as wilful blindness.**
>
> Chris Sidoti, expert member of the IIFFMM[256]

When seeking to assess the extent of Meta's role in respect of the 2017 atrocities against the Rohingya, an important factor is whether it was foreseeable that Meta risked contributing to human rights harms. According to international human rights standards, if a company knows, or should know, that it risks contributing to human rights harms, then it has a responsibility to take effective action to prevent and mitigate any human rights harms and risks. Accordingly, this chapter surveys the various warnings and interventions which Meta received in relation to Myanmar between 2013-2017, and then examines the role that the company's culture may have played in exacerbating human rights risks.

Amnesty International has interviewed several civil society activists who were involved in trying to convince Meta to take urgent action to avoid its platform contributing to mass atrocities in Myanmar. The organization has also analysed relevant United Nations and civil society reports and other publicly available materials which also warned of Meta's role in the increasing risk of an outbreak of mass anti-Muslim violence. These warnings are outlined in detailed below.

[256] Interview by voice call with Chris Sidoti, 13 April 2022.

7.1 WARNINGS RECEIVED BY META THAT IT RISKED CONTRIBUTING TO MASS VIOLENCE IN MYANMAR

As noted in Section 5.1 above, the Facebook platform played a prominent role in outbreaks of anti-Muslim violence in Myanmar both in 2012 and 2014 – and was even temporarily blocked by the authorities in 2014 in an effort to quell rising tensions and violence in Mandalay. In addition to these major incidents, Meta received multiple direct communications from experts and civil society activists warning it of the risk that it could contribute to a genocide or another form of mass violence unless it took urgent action to address emerging issues related to the use of the platform in Myanmar.

Collectively, these warnings paint a damning picture of the extent of Meta's repeated failure to act. These explicit warnings to Meta began as early as 2012, and continued through 2013, 2014, 2015, 2016 and 2017, as outlined in detail below.[257]

7.1.1 TIMELINE OF INTERVENTIONS

In November 2012, Htaike Htaike Aung, then programme director of Myanmar ICT for Development Organisation (MIDO), a local NGO, raised the issue of "hate speech" on the Facebook platform with Meta's director of global public policy as part of a roundtable organized by Freedom House at the Internet Governance Forum in Baku, Azerbaijan.[258]

In October 2013, Htaike Htaike Aung reiterated her concerns at a Freedom House roundtable held at the Internet Governance Forum in Bali, Indonesia, which was attended by a number of Meta policy executives.[259]

In November 2013, academic and journalist Aela Callan was conducting a fellowship focusing on the issue of "hate speech" in Myanmar at Stanford University. She travelled to Menlo Park in late 2013 and secured a meeting with Meta's Vice-President for Global Communications and Public Policy. At the meeting, she "raised explicitly this hate speech problem in Myanmar and urged them to take it seriously".[260] However, according to David Madden, founder of the Myanmar-based Phandeeyar foundation, the Meta staff member "wasn't that interested in the hate speech problem" because Myanmar was "seen as a tremendous market opportunity, certainly in terms of user growth".[261]

In March 2014, Htaike Htaike Aung, accompanied by Aela Callan, had a meeting with staff from Meta's 'compassion team' on the side of RightsCon 2014 and made another attempt at convincing Meta to take action to address the escalating risks in Myanmar.[262]

Also in March 2014, six Meta employees joined a call with academics and civil society experts to discuss concerns related to the spread of "hate speech" on the platform. Yangon-based academic Matt Schissler was invited by Harvard academic Susan Benesch to join the call and present on Myanmar. According to one account, Schissler "gave a stark recounting of how Facebook was hosting dangerous Islamophobia" and detailed 'the dehumanizing and disturbing language people were using in posts and the doctored photos and misinformation being spread widely".[263]

In April 2014, Meta's compassion team held a virtual meeting with staff from MIDO, Aela Callan, Susan Benesch and Matt Schissler, where they discussed cooperating on the localization of Meta's reporting tool for Facebook. A Facebook group was set up to support further consultation and engagement. Over the following

[257] Note: this timeline does not purport to include every relevant warning and intervention communicated to Meta in relation to emerging risks in Myanmar. Amnesty International is aware of several additional interventions which are not detailed in this timeline.

[258] Briefing deck by Victoire Rio and Htaike Htaike Aung, 13 June 2022. Upcoming release on Myanmar Internet Project, myanmarinternet.info.

[259] Briefing deck by Victoire Rio and Htaike Htaike Aung, 13 June 2022. Upcoming release on myanmarinternet.info.

[260] PBS Frontline, *The Facebook Dilemma,* 19 June 2018, pbs.org/wgbh/frontline/interview/david-madden

[261] PBS Frontline, *The Facebook Dilemma,* 19 June 2018, pbs.org/wgbh/frontline/interview/david-madden

[262] Briefing deck by Victoire Rio and Htaike Htaike Aung, 13 June 2022. Upcoming release on Myanmar Internet Project, myanmarinternet.info.

[263] Sheera Frenkel & Cecilia Kang, *An Ugly Truth: Inside Facebook's Battle for Domination*, 2021, p. 189.

months, it came to be used as a key channel to alert the company to emerging concerns, as activists could find no other way of raising emergency situations.[264]

In the first week of July 2014, tensions were rising in Mandalay as rumours were spreading on Facebook blaming two Muslim men for the rape of a young Buddhist woman. According to one report, "[i]n the days leading up to the riots, NGO workers tried to warn the company in the private Facebook group, but they hadn't heard back from anyone. Now people were getting killed, and there was still no response".[265] The Myanmar authorities blocked Facebook on the third day of the riots. According to Matt Schissler, when he reached out to a Meta contact about this, he received an immediate response – in contrast to the non-response to earlier concerns that people were being killed.[266]

In June 2014, Al Jazeera published an article entitled, "Facebook in Myanmar: Amplifying hate speech?". A Meta spokesperson contributed to the article.[267]

In July 2014, following the Mandalay riots and the authorities' move to block the platform, a Meta representative flew into Myanmar to take part in a panel discussion at the invitation of the Myanmar government.[268] This was the company's first official visit to Myanmar. The representative also conducted meetings with civil society activists during her visit.[269] Following these meetings, Meta took some minor steps aimed at responding to the concerns raised by civil society – namely by issuing a Burmese-language translation of its community standards, and by launching a 'flower speech' sticker, which was intended to be used by users in Myanmar to counter content that incited violence and discrimination.[270] See Section 6.3 on the unintended effects of these stickers on the amplification of harmful content.

On 17 March 2015, Matt Schissler travelled to California and made a presentation at Meta headquarters aimed at raising awareness within the company of the risk that Meta could contribute to anti-Muslim violence in Myanmar. Schissler shared "a PowerPoint presentation that documented the seriousness of what was happening in Myanmar: hate speech on Facebook was leading to real-world violence in the country, and it was getting people killed".[271] Schissler had a lunch meeting with a smaller group of Meta staff after the presentation. During the lunch, one employee asked Schissler if he thought Facebook could contribute to a genocide in Myanmar, to which he responded that yes, it was a real possibility.[272]

In May 2015, David Madden, the founder of the Phandeeyar foundation, made a presentation at Meta headquarters in Menlo Park, the purpose of which was "to try to help people understand what was going on, help Facebook decision makers understand what was going on in Myanmar at the time, and just how dangerous the situation was".[273] Madden told Amnesty International that "those of us who were working on these issues in Myanmar had a sense that people in Facebook didn't appreciate the nature of the political situation in the country".[274] During this presentation, attended by relevant Meta staff in person and via videoconference, Madden cited "examples of the kinds of content that we had already seen on Facebook that was being used to exacerbate divisions between communities and to ratchet up the temperature".[275] Madden provided explicit warnings that Meta risked contributing to mass violence in Myanmar:

> **"I drew the analogy with what had happened in Rwanda. There had been genocide in Rwanda, and radios had played a really key role in the execution of this genocide in Rwanda. And my concern was that Facebook would play a similar role in Myanmar, meaning it would be the platform through which hate speech was spread and incitements to violence were made.**

[264] Briefing deck by Victoire Rio and Htaike Htaike Aung, 13 June 2022. Upcoming release on Myanmar Internet Project, myanmarinternet.info.

[265] Sheera Frenkel & Cecilia Kang, *An Ugly Truth: Inside Facebook's Battle for Domination*, 2021, p. 190.

[266] Sheera Frenkel & Cecilia Kang, *An Ugly Truth: Inside Facebook's Battle for Domination,* 2021, p. 190.

[267] Hereward Holland, "Facebook in Myanmar: Amplifying hate speech?", June 2014, *Al Jazeera*, https://www.aljazeera.com/features/2014/6/14/facebook-in-myanmar-amplifying-hate-speech

[268] See: Aung Kyaw Zaw, "2-1 U Htin Aung Hlaing Discussion Partner by U Chit Htun Phe Alexander Long Mla Garlick John Karr", YouTube, youtube.com/watch?v=R32WQNbKyk4&ab_channel=AungKyawZaw

[269] Briefing deck by Victoire Rio and Htaike Htaike Aung, 13 June 2022. Upcoming release on Myanmar Internet Project, myanmarinternet.info.

[270] See: "Support Panzagar: End Hate Speech with Flower Speech", Facebook Page, facebook.com/supportflowerspeech

[271] Sheera Frenkel & Cecilia Kang, *An Ugly Truth: Inside Facebook's Battle for Domination*, 2021 p. 193.

[272] Sheera Frenkel & Cecilia Kang, *An Ugly Truth: Inside Facebook's Battle for Domination*, 2021, p. 193.

[273] PBS Frontline, *The Facebook Dilemma,* 19 June 2018, pbs.org/wgbh/frontline/interview/david-madden

[274] Interview by video call with David Madden, 09 June 2022.

[275] Interview by video call with David Madden, 09 June 2022.

> **And so, I said very clearly to them that Facebook runs the risk of being in Myanmar what radios were in Rwanda. I said that very clearly; I said it very explicitly. It wasn't the last time that I said it. I said it on many occasions after that. But I think that was the first time that I had said it to them."[276]**

In September 2015, a senior Meta representative travelled to Myanmar to launch the Burmese-language version of Facebook's Community Standards. During a meeting between the Meta representative and a range of local civil society groups, several groups raised concerns that Facebook's community standards were not being adequately enforced in the country.[277]

In 2016, serious violence erupted in Northern Rakhine state as the Myanmar security forces engaged in a brutal crackdown against the Rohingya population after the ARSA militant group attacked border posts. The violence foreshadowed the events to come in 2017.[278] A senior Meta representative made at least three visits to Myanmar in 2016. In one meeting in November 2016 between the Meta representative and local digital rights activists, local activists delivered a "stark warning" about the risks that Meta could contribute to mass violence in Myanmar.[279]

In January 2017, David Madden and Victoire Rio remotely joined a meeting between Meta representatives and another digital rights activist taking place in Menlo Park.[280] David Madden recounted the briefing to PBS Frontline:

> **"We were genuinely worried about where things might go from there, and the situation on Facebook was even worse because what was really apparent by now was just how rife the hate speech problem was, and importantly, just how inadequate Facebook's response was. So we were very clear at that meeting that their systems just didn't work. The processes that they had in place to try to identify and pull down problematic content, they just weren't working. And we were deeply concerned about where this was going to go, and the possibility that something even worse was going to happen imminently…**
>
> **So we were very prescriptive and very clear at that meeting. At that meeting I reiterated this point that there was a real risk that Facebook would be in Myanmar what the radios had been in Rwanda, and I was really clear about that."[281]**

In June 2017, a Meta delegation travelled to Myanmar and met with local digital rights groups including Phandeeyar and the Myanmar ICT for Development Organization (MIDO). Victoire Rio, who attended two meetings with Meta staff during that trip, told Amnesty International that she "raised the fact that we needed a much better escalation mechanism, and that policies really needed to be contextualized to account for risks".[282]

Despite receiving these repeated direct warnings and pleas for action, Meta failed, again and again, to take action which could have mitigated its adverse human rights impacts in Myanmar, as detailed further in Section 8.2 below.

7.2 META'S KNOWLEDGE OF THE RISKS INHERENT IN ITS ALGORITHMS

There is a substantial body of evidence which suggests that Meta was aware, or at least should have been aware, that its content-shaping algorithms risked contributing to adverse human rights impacts in Myanmar since as early as 2011. In a best-selling 2011 book, author and activist Eli Pariser outlined the risks inherent in the algorithms used on social media platforms, including Facebook:

[276] PBS Frontline, *The Facebook Dilemma,* 19 June 2018, pbs.org/wgbh/frontline/interview/david-madden

[277] Briefing deck by Victoire Rio and Htaike Htaike Aung, 13 June 2022. Upcoming release on Myanmar Internet Project, myanmarinternet.info.

[278] See: Amnesty International, '"We are at breaking point" Rohingya: persecuted in Myanmar, neglected in Bangladesh', 2016, ASA 16/5362/2016, 19 December 2016, amnesty.org/en/documents/asa16/5362/2016/en

[279] Interview by voice call with Victoire Rio, 13 June 2022.

[280] Interview by voice call with Victoire Rio, 13 June 2022.

[281] PBS Frontline, *The Facebook Dilemma,* 19 June 2018, pbs.org/wgbh/frontline/interview/david-madden

[282] Interview by video call with Victoire Rio, 12 April 2022.

> **"Personalization filters serve a kind of invisible autopropaganda, indoctrinating us with our own ideas, amplifying our desire for things that are familiar and leaving us oblivious to the dangers lurking in the dark territory of the unknown."[283]**

Studies conducted by Meta researchers have repeatedly shown how the platform's content-shaping algorithms can negatively impact users' thoughts and behaviours. In 2012, Meta researchers undertook an experiment on 689,003 Facebook users which sought to assess the impact of curated News Feed content on users' behaviour and mood. The research found that by algorithmically tweaking the content seen by Facebook users, "emotional states can be transferred to others via emotional contagion, leading people to experience the same emotions without their awareness".[284] Meta researchers again studied the impact of Meta's algorithms on political polarization in 2015.[285] The study found that "Facebook's newsfeed algorithm decreases ideologically diverse, cross-cutting content people see from their social networks on Facebook by a measurable amount".[286]

As noted in Section 6.2.4, internal Meta research from 2016 found "extremist content thriving in more than one-third of large German political groups on the platform", and, significantly, it also recognized that Meta's content-shaping algorithms were responsible for the growth of these groups.[287] The report noted:

> **The 2016 presentation stated that "64% of all extremist group joins are due to our recommendation tools" and that most of the activity came from the platform's "Groups You Should Join" and "Discover" algorithms: "*Our recommendation systems grow the problem* [emphasis added]."[288]**

These studies should have made Meta aware that its content-shaping algorithms posed a significant risk to human rights, especially in sensitive and conflict-affected contexts – long before the 2017 atrocities in Northern Rakhine State began.

Though they mostly originate after 2017, the documents contained in the Facebook Papers reveal in greater detail than ever before the extent of Meta's detailed knowledge of the harms associated with its business model and engagement-centric algorithms. They also highlight how the company has continued to pursue engagement at all costs even in the face of such detailed evidence. In one document contained in the Facebook Papers, entitled '*What is Collateral Damage?*', dated 12 August 2019, an employee comments:

> **"We have evidence from a variety of sources that hate speech, divisive political speech, and misinformation on Facebook and the family of apps are affecting societies around the world. We also have compelling evidence that our core product mechanics, such as virality, recommendations, and optimizing for engagement, are a significant part of why these types of speech flourish on the platform. Therefore, if integrity takes a hands-off stance for these problems, whether for technical (precision) or philosophical reasons, then the net result is that Facebook, taken as a whole, will be actively (if not necessarily consciously) promoting these types of activities. The mechanics of our platform are not neutral."[289]**

In another internal document entitled 'Facebook and Responsibility', dated March 2020, the author offers a more detailed explanation of the extent of Meta's responsibility for the content it shows to users on the platform:

[283] Eli Pariser, "The Filter Bubble: What The Internet Is Hiding From You", 2011, p.13.

[284] Adam D.I. Kramer et al, "Experimental evidence of massive-scale emotional contagion through social networks", *Psychological and Cognitive Sciences*, 2 June 2014, pnas.org/doi/full/10.1073/pnas.1320040111. See also: Robinson Meyer, "Everything We Know About Facebook's Secret Mood-Manipulation Experiment", *The Atlantic*, 28 June 2014, theatlantic.com/technology/archive/2014/06/everything-we-know-about-facebooks-secret-mood-manipulation-experiment/373648

[285] Zeynep Tufekci, "How Facebook's Algorithm Suppresses Content Diversity (Modestly) and How the Newsfeed Rules Your Clicks", 7 May 2015, *Medium*, medium.com/message/how-facebook-s-algorithm-suppresses-content-diversity-modestly-how-the-newsfeed-rules-the-clicks-b5f8a4bb7bab

[286] Zeynep Tufekci, "How Facebook's Algorithm Suppresses Content Diversity (Modestly) and How the Newsfeed Rules Your Clicks", 7 May 2015, *Medium*, medium.com/message/how-facebook-s-algorithm-suppresses-content-diversity-modestly-how-the-newsfeed-rules-the-clicks-b5f8a4bb7bab

[287] The Wall Street Journal, "Facebook Executives Shut Down Efforts to Make the Site Less Divisive", 26 May 2020, wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499

[288] The Wall Street Journal, "Facebook Executives Shut Down Efforts to Make the Site Less Divisive", 26 May 2020, wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499

[289] The Facebook Papers, "What is Collateral Damage?", 12 August 2019, comment on p. 34.

> **"Facebook will generally have some responsibility for exposing users to harmful content whenever users saw that content as a result of Facebook's actions … from creating surfaces that deliver content to developing algorithms that make decisions about where content appears. If those surfaces or algorithms could have been designed differently, then Facebook has taken some action. The more active Facebook is in determining whether a user is exposed to harmful content, the most responsibility we have."[290]**

These admissions of the adverse human rights impacts associated with Meta's content-shaping algorithms clearly run counter to the company's refusal to remediate affected communities, as explored further in Chapter 9, below.

Despite the litany of warnings that Meta received between 2012 and 2019 that it risked contributing to an outbreak of mass violence, and despite the fact that Meta did in fact contribute to mass violence against the Rohingya (as analysed further in Chapter 8, below), the Facebook Papers show that the company has continued to deploy the same dangerous content-shaping algorithms in the pursuit of maximum profit. As a result, in the absence of significant action to prevent and mitigate further harm, there is a real risk the company may contribute to an outbreak of mass violence again in other sensitive contexts.

7.3 'MOVING FAST AND BREAKING THINGS': HOW META'S COMPANY CULTURE EXACERBATED HUMAN RIGHTS RISKS

In September 2021, Facebook rebranded as Meta and announced its plans to invest heavily in research to ensure its metaverse products "are developed responsibly."[291] At the time, Meta's Vice President of Global Affairs, Nick Clegg, stated that this new approach was "almost the opposite of that now long-abandoned slogan of 'move fast and break things'."[292] The statement arguably represents a tacit acknowledgement that Meta's prior approach was 'the opposite' of responsible conduct. In September 2021, after the publication of the Facebook Papers, Meta published an article entitled, "Our Progress Addressing Challenges and Innovating Responsibly". The article admitted:

> **"In the past, we didn't address safety and security challenges early enough in the product development process. Instead, we made improvements reactively in response to a specific abuse."[293]**

Meta's original "move fast and break things" motto reflected the company's overall business ethos.[294] The approach meant that Meta staff were encouraged to innovate in a way that would maximize rapid growth, and problems that arose were to be understood and resolved later.[295] In 2014, Mark Zuckerberg announced a departure from "move fast and break things".[296] However, as recently as 2021 it was reported that: "Every engineer at the company says, 'This is still our motto.'"[297] According to one former Meta employee, the "move fast and break things" slogan was "still used as a holdover in some spaces" within the company.[298]

[290] The Facebook Papers, "Facebook and Responsibility", 09 March 2020, p.3, documentcloud.org/documents/21594152-tier2_rank_other_0320

[291] Andrew Bosworth and Nick Clegg, "Building the Metaverse Responsibly", 27 September 2021, fb.com/news/2021/09/building-the-metaverse-responsibly/, referenced in Axios, "Facebook: Metaverse won't move fast and break things", 28 September, 2021, axios.com/2021/09/28/facebook-metaverse-research-nick-clegg

[292] Axios, "Facebook: Metaverse won't move fast and break things", 28 September 2021, axios.com/2021/09/28/facebook-metaverse-research-nick-clegg

[293] Meta, "Our Progress Addressing Challenges and Innovating Responsibly", 21 September 2021, fb.com/news/2021/09/our-progress-addressing-challenges-and-innovating-responsibly

[294] Business Insider, "Mark Zuckerberg On Innovation", 1 October 2009, businessinsider.com/mark-zuckerberg-innovation-2009-10?r=US&IR=T

[295] Jonathan Taplin, *Move Fast and Break Things: How Facebook, Google, and Amazon Cornered Culture and Undermined Democracy,* 2017.

[296] Wired, "Mark Zuckerberg on Facebook's Future From Virtual Reality to Anonymity", 30 April 2014, wired.com/2014/04/zuckerberg-f8-interview. Zuckerberg described the rationale for the change in approach and motto as: "[so that] people can rely on us as a critical infrastructure for building all of their apps across every mobile platform".

[297] Forcepoint, *F**k It, Ship It! With Sheera Frenkel,* 7 September 2021, forcepoint.com/resources/podcast/f--k-it-ship-it-sheera-frenkel

[298] Interview by video call with a former Meta employee, 27 April 2022.

Previous investigations into Meta's internal culture portray a culture which incentivized and rewarded high-risk behaviours, [299] while staff who sought to raise the alarm about human rights risks, and address them, were faced with major structural barriers. [300]

Meta's propensity for human rights risk and its consistent failure to proactively address its responsibilities represent the symptoms of a business model which inherently conflicts with international human rights standards. As Amnesty International has previously reported, Meta's business model is predicated on targeted advertising enabled by the extraction and accumulation of vast amounts of data about people. To increase its revenue from advertisers, Meta needs to continuously expand its data vault and refine its predictive algorithms to maximize engagement.[301]

In September 2021, when Meta recognized the failings of its prior approach to "innovating" (grounded in the "move fast and break things" motto) following the publication of the Facebook Papers, the company also claimed:

> **"… but we have fundamentally changed that approach. Today, we embed teams focusing specifically on safety and security issues directly into product development teams, allowing us to address these issues during our product development process, not after it. Products also have to go through an Integrity Review process, similar to the Privacy Review process, so we can anticipate potential abuses and build in ways to mitigate them."** [302]

Meta's original "move fast and break things" motto entailed clear tensions with the company's responsibility to undertake proactive and comprehensive human rights due diligence in respect of all of its business activities. Meta's recent statements claiming that is has adopted a "fundamentally changed" approach to "safety and security issues" did not explicitly refer to the role of human rights due diligence. It remains to be seen what impact, if any, this change of approach will have on Meta's human rights footprint. And even if it does entail substantive improvements, for many affected communities across the world – including the Rohingya – the changes have come far too late.

On 20 May 2022, Amnesty International wrote to Meta and asked what measures the company took prior to 2017 to prevent its algorithmic systems from being used to amplify advocacy of hatred against the Rohingya. Meta responded that "Meta's investments in Myanmar in response to the events in 2017 have been significant", but added that the company could not provide information concerning the period leading up to 2017 because the company is "currently engaged in litigation proceedings in relation to related matters".[303]

In March 2021, Meta adopted a human rights policy for the first time.[304] In the policy, the company committed to "reporting annually on how we are addressing our human rights impacts".[305] Meta published its first such human rights report in July 2022.[306] However, the report failed to acknowledge human rights risks and impacts associated with the company's business model, and made no mention of the company's use of content-shaping algorithms.[307]

Meta's company culture and its data-hungry business model incentivize the company to rapidly expand its operations across the globe and into local contexts that entail serious human rights risks, including settings

[299] See, for example: Karen Hao, "How Facebook got addicted to spreading misinformation", *MIT Technology Review*, 11 March 2021 technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation; WSJ: "Facebook Executives Shut Down Efforts to Make the Site Less Divisive", 26 May 2020, wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499

[300] Julia Carrie Wong, "The Facebook Loophole: How Facebook let fake engagement distort global politics: a whistleblower's account", *The Guardian*, 12 April 2021 theguardian.com/technology/2021/apr/12/facebook-fake-engagement-whistleblower-sophie-zhang

[301] Amnesty International, *Surveillance Giants: How the business model of Google and Facebook threatens human rights* (Index: POL 30/1404/2019), 21 November 2019, amnesty.org/en/documents/pol30/1404/2019/en, p. 12

[302] Meta, "Our Progress Addressing Challenges and Innovating Responsibly", 21 September 2021, about.fb.com/news/2021/09/our-progress-addressing-challenges-and-innovating-responsibly

[303] Meta, Letter to Amnesty International, 31 May 2022 (on file with Amnesty International).

[304] Meta, "Corporate Human Rights Policy, 16 March 2021, about.fb.com/wp-content/uploads/2021/04/Facebooks-Corporate-Human-Rights-Policy.pdf

[305] Meta, "Human Rights Report 2020-2021", about.fb.com/wp-content/uploads/2022/07/Meta_Human-Rights-Report-July-2022.pdf

[306] Meta, "Human Rights Report 2020-2021", about.fb.com/wp-content/uploads/2022/07/Meta_Human-Rights-Report-July-2022.pdf

[307] See: Amnesty International, "Meta's Human Rights Report ignores the real threat the company poses to human rights worldwide", 22 July 2022, amnesty.org/en/latest/campaigns/2022/07/metas-human-rights-report-ignores-the-real-threat-the-company-poses-to-human-rights-worldwide

affected by conflict.[308] As noted in Chapter 3, the risks of conflict and violence breaking out in Myanmar should have been abundantly clear prior to Meta's market entry into Myanmar. These risks should have become even clearer during the company's early years operating in in the country. As outlined in Section 4.1, conflicted-affected settings require a higher-than-usual standard of human rights due diligence to be conducted, as such settings carry especially severe risks that the company may cause or contribute to human rights harms. However, as detailed further in Section 8.2, below, Meta's company culture and business model meant that the company failed to incorporate appropriate human rights due diligence measures at the time of its market entry and during the early years of its operations in Myanmar.

[308] See: Amnesty International, *Surveillance Giants* (previously cited), p. 12.

8. META'S RESPONSIBILITY FOR HUMAN RIGHTS HARMS SUFFERED BY THE ROHINGYA

In this section, Meta's role in the human rights abuses suffered by the Rohingya in the context of the 2017 atrocities is analysed based on the evidence and human rights standards laid out in the preceding sections. As established under the UN Guiding Principles, the key question to answer when seeking to assess a company's responsibility to provide an effective remedy to an affected community is whether the company's actions reached the threshold of ***causing*** or ***contributing*** to the adverse human rights impact. This analysis seeks to ascertain whether Meta caused, contributed, or was directly linked to the adverse human rights impacts suffered by the Rohingya.

When assessing whether Meta fulfilled its responsibility to respect human rights in the context of its business activities in Myanmar, there are two adverse human rights impacts that are particularly important: the discrimination towards the Rohingya created by the dissemination of harmful speech including advocacy of hatred (as detailed in Section 5.2), and secondly, the violence that the dissemination of this content contributed to, which had knock-on impacts on the rights to life and freedom from torture; and the rights to education, to adequate housing and an adequate standard of living, and the right to food (as detailed in Sections 3.2 and 5.3).

The UN Guiding Principles and the OECD Guidelines provide guidance on how to assess whether a company should provide an effective remedy when they are implicated in an adverse human rights impact.[309] This guidance (outlined in detail in Chapter 4) is incorporated throughout the following assessment of Meta's role in respect of the human rights harms suffered by the Rohingya.

[309] BSR, White Paper: Seven Questions to Help Determine When a Company Should Remedy Human Rights Harm under the UN Guiding Principles, 2021.

8.1 META'S ROLE IN HUMAN RIGHTS HARMS

In seeking to assess whether a company has reached the threshold of 'contributing' to an adverse human rights impact, a variety of factors will be taken into account, and will vary depending on the context and the business activity. The author of the UN Guiding Principles, Professor John Ruggie, has explained:

> **"There is a continuum between contribution and linkage. A variety of factors can determine where on that continuum a particular instance may sit [including] the extent to which a business enabled, encouraged, or motivated human rights harm by another; the *extent to which it could or should have known about such harm*; and the *quality of any mitigating steps it has taken to address it* [emphasis added]."[310]**

The OECD Due Diligence Guidance (2018) further explains:

> **An enterprise "contributes to" an impact if its activities, in combination with the activities of other entities cause the impact, or if the activities of the enterprise cause, facilitate or incentivise another entity to cause an adverse impact.[311]**

The Facebook platform's core functionality – as a social media platform enabling mass dissemination of communications by users – enabled other parties to disseminate anti-Rohingya content, including advocacy of hatred that incited violence and discrimination. This was especially true in Myanmar, where Meta was so completely dominant that its Facebook platform was most people's primary news source and became synonymous with the internet.

Meta's business model relies on the invasive collection and analysis of vast amounts of data about people in the digital world, which can then be monetised for the purpose of targeted advertising. Meta's core product mechanics for Facebook – its content-shaping algorithms – are optimized to ensure that users engage with content on Facebook as much as possible and spend as much time as possible on the platform. The more engaged users are, and the more time they spend on the platform, the more advertising revenue Meta earns. Research has consistently shown that the result of Meta algorithms' optimization for maximum engagement is that the most inflammatory, divisive and harmful content is prioritized (See Section 6.2).

As outlined in Sections 6.1 and 6.2, above, Meta's use of content-shaping algorithms which were designed to maximize engagement, including its news feed and recommendations algorithms, facilitated the dissemination of advocacy of hatred against the Rohingya. By optimizing for engagement and amplifying the most inflammatory content, in a context where anti-Muslim and anti-Rohingya sentiment was rampant, Meta effectively delivered these messages on behalf of third parties to the exact audiences which were most likely to be *engaged* by this content – and by extension – motivated towards supporting or engaging in mass violence. As such, the evidence outlined in the previous sections clearly supports the contention that Meta, through the deployment of its content-shaping algorithms, enabled and facilitated the dissemination of anti-Rohingya content and advocacy of hatred.

Meta also created financial incentives for third parties to produce and share harmful content through its monetisation drive which began in 2015, and which was rolled out in Myanmar in 2016. As outlined in Section 6.4, above, Meta's 'Instant Article' feature and its associated monetization drive, along with the platform's engagement-centric algorithms, combined to create financial incentives for 'creators' to publish the most engaging, and therefore inflammatory, content they could.

In addition to profiting from the dissemination of inflammatory anti-Rohingya content because of its impact of Meta's engagement metrics, Meta hosted paid advertising on Facebook from the Myanmar military and prominent anti-Rohingya hate group Ma Ba Tha, which would have enabled the groups to micro-target their audiences.[312] Meta would have received direct payments from these groups or their representatives in order for these ads to have been shown. Although Amnesty International does not have direct proof of these

[310] John G Ruggie, "Comments on Thun Group of Banks Discussion Paper on the Implications of UN Guiding Principles 13 & 17 In a Corporate and Investment Banking Context", 21 February 2017, business-humanrights.org/sites/default/files/documents/Thun%20Final.pdf, cited in BSR, "White Paper: Seven Questions to Help Determine When a Company Should Remedy Human Rights Harm under the UN Guiding Principles", 7 January 2021, bsr.org/en/our-insights/report-view/seven-questions-to-help-determine-when-a-company-should-remedy-human-rights [emphasis added]

[311] OECD Due Diligence Guidance (previously cited), p. 70 [emphasis added].

[312] Victoire Rio, "The Role of Social Media in Fomenting Violence: Myanmar".

transactions, it should be noted that this criterion (direct financial benefit) is not required in order for Meta to reach the "contribution" threshold, according to the UN Guiding Principles.[313]

As such, it can be concluded that Meta **enabled and facilitated** the dissemination of harmful content including advocacy of hatred against the Rohingya insofar as its core features and services (specifically its news feed, ranking, and recommendation algorithms) actively amplified, promoted, and delivered these messages to the target audiences who were most likely to act upon them (as detailed in Sections 6.2 and 6.3). It also **incentivized** these harms through its monetization drive and the deployment of its 'Instant Article' feature.

8.2 META'S DUE DILIGENCE FAILURES IN MYANMAR

If a company has actual knowledge that a harm is likely to occur but fails to take action to prevent it, *or* if it should have known that harm was likely to occur had it undertaken reasonable due diligence, the likelihood that the company will be found to have contributed to the harm is higher. OECD guidance states that the "extent to which an enterprise could or should have known about the adverse impact or potential from adverse impact, i.e., the degree of foreseeability" is a key factor in assessing whether a company reached the threshold of 'contribution'.[314]

There are two key components in the assessment of whether Meta knew or should have known that it risked contributing to human rights harms in Myanmar: 1) whether and when Meta knew or should have known that its content-shaping algorithms, such as its news feed and recommendations algorithms, disproportionately favoured inflammatory and dehumanizing content; and, 2) whether the company knew or should have known about the risks which were particular to its business operations and market entry in Myanmar. Both components are assessed in turn below. This section concludes with an assessment of Meta's overall fulfilment of its responsibility to conduct adequate human rights due diligence in Myanmar.

8.2.1 META'S KNOWLEDGE OF RISK OF ALGORITHMIC HARM

The evidence, as laid down in Section 7.2 above, reveals that Meta had *actual* knowledge of certain risks associated with its algorithms as early as 2011. This evidence includes multiple internal studies conducted by the company dating from 2012, 2015, and 2016 (as detailed in Section 7.2), in addition to prominent studies by independent researchers and authors.

These internal studies were not framed explicitly around human rights impacts. Nonetheless, there is sufficient evidence to conclude that the company, at the very least, *should* have been aware of how these risks could contribute to adverse human rights impacts, had it engaged in adequate human rights due diligence. However, research conducted by Amnesty International and others (see Section 7.3, above), suggests that Facebook's company culture and business model resulted in a dereliction of its responsibility to engage in ongoing and proactive human rights due diligence in respect of its content-shaping algorithms. Had it conducted appropriate human rights due diligence, and embedded these practices into company culture, Meta could have proactively identified and mitigated human rights risks in Myanmar.

In May 2022, Amnesty International asked Meta when it became aware of risks that the company's "core product mechanics", including its ranking and recommendation algorithms, could promote and amplify advocacy of hatred and other harmful content on its platform. In its response, Meta did not provide an answer to this question.[315]

8.2.2 META'S KNOWLEDGE OF MYANMAR-SPECIFIC RISKS

As outlined in Chapter 3, the Myanmar context when Meta first entered the market in the early 2010s was characterized by human rights risks and amounted to a conflict-affected setting. Had Meta conducted appropriate proactive human rights due diligence before deploying its content-shaping algorithms in Myanmar, it is likely that it would have identified a risk that it could amplify content advocating hatred against the Rohingya. As noted in Chapter 4, in conflict-affected settings such as Myanmar, Meta had an

[313] UN Guiding Principles Interpretative Guide, pp. 5 and 79.

[314] OECD Due Diligence Guidance (2018), p. 70.

[315] Meta, Letter to Amnesty International, 31 May 2022, on file with Amnesty International.

enhanced responsibility to conduct comprehensive due diligence through processes that were capable of capturing unpredictable as well as rapid changes in risks.[316]

As outlined in detail in Section 7.1, following Meta's market entry into Myanmar, the company received repeated warnings from civil society activists that it risked contributing to an outbreak of mass violence, and repeatedly failed to act upon them, including several specific instances prior to 2017 where the company's platform was used to spread harmful content including advocacy of hatred against the Rohingya, resulting in ethnic violence. These repeated warnings and interventions mean that Meta either knew or should have known about the potential human rights harms which it risked contributing to in Myanmar.

As noted in Chapter 4, Meta had a responsibility to engage in ongoing human rights due diligence in respect of its business activities in Myanmar, encompassing a responsibility to draw upon "internal and/or independent external human rights expertise" and engage in "meaningful consultation with potentially affected groups and other relevant stakeholders".[317] However, Meta's engagement with local civil society and human rights experts in Myanmar was generally reactive in nature, and these stakeholders consistently described the inadequacy of Meta's engagement prior to 2017 (as detailed in Section 7.1).

Amnesty International asked Meta whether the company carried out human rights due diligence related to the company's presence in Myanmar between 2011 and 2017. Meta did not provide any information because the company is "currently engaged in litigation proceedings in relation to related matters".[318]

8.3 META'S FAILURE TO EFFECTIVELY MITIGATE HUMAN RIGHTS RISKS

To fulfil its responsibility to respect human rights, Meta had a responsibility to "[a]void causing or contributing to adverse human rights impacts through their own activities" and to "seek to prevent or mitigate adverse human rights impacts that are directly linked to their operations, products or services by their business relationships, even if they have not contributed to those impacts."[319]

Meta persistently failed to take adequate action aimed at preventing harm and mitigating human rights risks in response to repeated warnings from civil society and human rights experts between 2012 and 2017. The minor steps it took in 2014 – such as the launch of its 'flower speech sticker pack' and translating its community standards into Burmese – did not seek to address the amplification of harmful content, and were ultimately negligible in their impact. As noted in Section 5.3, the flower speech stickers may have even further amplified harmful content, including advocacy of hatred against the Rohingya, on the Facebook platform.

There are numerous actions which Meta could have taken to prevent the dissemination and amplification of advocacy of hatred against the Rohingya and other anti-Rohingya content. In recent years, Meta has developed so-called 'Break the Glass' (BTG) measures intended for deployment in critical emergencies. BTG measures – referred to as 'levers' in internal company documents – effectively switch off, or at least significantly reduce, the amplifying power of Meta's engagement-centric algorithms.[320] The existence of these measures appears to be based on a tacit acknowledgement that Meta's content-shaping algorithms can promote divisiveness and inflammatory content. The company has deployed BTG measures in the U.S. on at least two occasions – the 2020 Presidential elections, and the 6 January 2021 riots in Capitol Hill. [321]

BTG measures are short-term by nature; once the peak of a given crisis is perceived to have passed, the levers are switched back on and normal service for Meta's algorithms resumes. It is highly questionable

[316] Farhad Manjoo, "Facebook Use Polarizing? Site Begs to Differ", 7 May 2015, *The New York Times*, nytimes.com/2015/05/08/technology/facebook-study-disputes-theory-of-political-polarization-among-users

[317] UN Guiding Principles, Principle 18.

[318] Meta, Letter to Amnesty International, 31 May 2022, on file with Amnesty International.

[319] UN Office of the High Commissioner for Human Rights (UNOHCHR), 'Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework', 2011, UN Doc HR/PUB/11/04, Principle 13 including Commentary, ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf

[320] See: Donie O'Sullivan, "Not stopping 'Stop the Steal:' Facebook Papers paint damning picture of company's role in insurrection", CNN, 24 October 2021, edition.cnn.com/2021/10/22/business/january-6-insurrection-facebook-papers/index.html

[321] See: Donie O'Sullivan, "Not stopping 'Stop the Steal:' Facebook Papers paint damning picture of company's role in insurrection", CNN, 24 October 2021, edition.cnn.com/2021/10/22/business/january-6-insurrection-facebook-papers/index.html

whether such temporary measures would be sufficient to prevent and mitigate harms which Meta's content-shaping algorithms have helped to enable over a long-term period.

The Facebook Papers reveal that Meta has deployed BTG measures in Myanmar on at least one occasion. In the case (referenced in Section 5.3, above) wherein Meta found that 70% of the views gained by a video by hate figure U Wirathu had been actively chained to by Meta's algorithms, the company did launch "a BTG measure to stop video chaining" in Myanmar in response.[322]

However, Amnesty International could find no evidence of Meta deploying any similar measures – aimed at reducing algorithmic harms – in Myanmar ahead of the 2017 violence perpetrated against the Rohingya. Despite the repeated warnings and requests which the company received from civil society, Meta continued to deploy its content-shaping algorithms which were actively amplifying harmful content in Myanmar. Amnesty International wrote to Meta and asked what measures the company took prior to 2017 to prevent its algorithmic systems from being used to amplify advocacy of hatred against the Rohingya. Meta responded that the company could not provide information concerning the period leading up to 2017 because the company is "currently engaged in litigation proceedings in relation to related matters".[323]

In addition to the possibility of using its algorithmic levers to reduce the direct amplification of harmful content, Meta could have taken far more comprehensive action in respect of content moderation. As Section 5.4, above, details, Meta drastically underinvested in its content moderation for Myanmar, resulting in a situation where most reports of violating content either went ignored or were subject to patently incorrect determinations that the reported content did not violate Facebook's community standards.

Even a well-resourced approach to content moderation, in isolation, would likely not have sufficed to prevent and mitigate these algorithmic harms. This is because content moderation fails to address the root cause of Meta's algorithmic amplification of harmful content (See Section 6.1 for further analysis). Effective content moderation can nevertheless be a relevant mitigation tactic when it is combined with other measures that seek to directly prevent and mitigate algorithmic amplification. Ultimately, however, Meta's content moderation operations in Myanmar ahead of the 2017 violence were deeply inadequate. As such, Meta's use of content moderation cannot be deemed an effective mitigation measure ahead of the 2017 atrocities in Northern Rakhine State.

Amnesty international asked Meta what action the company took in Myanmar ahead of and during the crisis in 2017, as well as what steps the company took to address the risks of its "core product mechanics" amplifying advocacy of hatred and other harmful content. Meta did not provide any information because the company is "currently engaged in litigation proceedings in relation to related matters".[324]

8.3.1 AN "*INESCAPABLE*" CONCLUSION: META SUBSTANTIALLY CONTRIBUTED TO HUMAN RIGHTS HARMS SUFFERED BY THE ROHINGYA

> **"When can we say that a company like Facebook is 'contributing to' human rights harm?... Unwittingly getting even severely consequential cases wrong once or twice is one thing. But persistent refusal to substantially change what the company does to reduce its role in others' promotion of social strife and violence makes the attribution of 'contribution' inescapable."**
>
> John Ruggie, author of the UN Guiding Principles, 2018[325]

The independent human rights impact assessment for Myanmar commissioned by Meta did not overtly address whether the company had caused, contributed, or was directly linked to adverse human rights impacts against the Rohingya. This is despite the fact the HRIA was conducted in the immediate aftermath

[322] The Facebook Papers, "Video Time Spent in Myanmar", 12 November 2020, on file with Amnesty International, p.16

[323] Meta, letter to Amnesty International, 31 May 2022, on file with Amnesty International.

[324] Meta, letter to Amnesty International, 31 May 2022, on file with Amnesty International.

[325] John G. Ruggie, "Facebook in the Rest of the World," John F. Kennedy School of Government, Harvard University, November 15, 2018, https://media.business-humanrights.org/media/documents/files/documents/John_Ruggie_Facebook_15_Nov_2018.pdf, cited in Mark Latonero, and Aaina Agarwal, 'Human Rights Impact Assessments for AI: Learning from Facebook's Failure in Myanmar', Carr Center For Human Rights Policy Harvard Kennedy School, 2021.

of the 2017 atrocities in northern Rakhine State and triggered by the IIFFMM's findings in relations to the Facebook platform's "*significant"* role in the violence.[326] In its section on non-discrimination, the Rohingya are not mentioned at all.[327]

OECD Guidance states that, in order for a company's responsibility to provide remedy to affected communities to be triggered:

> **Contribution must be substantial, meaning that it does not include minor or trivial contributions.[328]**

Meta's contribution to adverse human rights impacts against the Rohingya was substantial in nature (not minor or trivial) because the core features of the Facebook platform (specifically its news feed, ranking, and recommendation algorithms) actively amplified, promoted, and delivered posts inciting advocacy of hatred and violence against the Rohingya to the target audiences who were most likely to act upon them. The effects of these actions for the Rohingya were especially severe because of the near-total dominance of the Facebook platform within Myanmar's online environment at the time. The substantial nature of this contribution is further underlined by the company's repeated failures to undertake appropriate human rights due diligence which could have effectively prevented or mitigated human rights risks, in addition to the severity of the consequences of the company's failures.

Amnesty International's systematic analysis of Meta's role in the respect of the serious human rights abuses suffered by the Rohingya of Northern Rakhine State in 2017 enables the following conclusions:

- Content that advocated hatred, violence, and discrimination against the Rohingya, as well as other dehumanizing and discriminatory anti-Rohingya, content was rife on the Facebook platform in the months and years leading up to the 2017 atrocities in Northern Rakhine State, with direct impacts on the right to equality and the right to be free from discrimination for the Rohingya.
- The prevalence of this content contributed to the offline violence perpetrated against the Rohingya, with knock-on impacts on the rights to life and freedom from torture, the right to education, to adequate housing and an adequate standard of living, and the right to food.
- Meta's content-shaping algorithms actively amplified and promoted divisive and inflammatory content, including content which dehumanized and discriminated against the Rohingya.
- Meta knew or should have known that it risked contributing to adverse human rights impacts against the Rohingya, having conducted multiple internal studies in relation to its algorithms and having received multiple warnings from civil society activists in relation to Myanmar-specific risks.
- Meta failed to engage in adequate human rights due diligence, which could and should have identified many of the risks that were present in relation to its operations in Myanmar. Meta also failed to enact effective and appropriate risk mitigation measures which could have prevented or mitigated these harms.
- Meta substantially contributed to adverse human rights impacts against the Rohingya, and the company has a responsibility under international human rights standards to remediate the harm suffered by this community.

[326] BSR, 2018, "Human Rights Impact Assessment: Facebook in Myanmar."

[327] BSR, 2018, "Human Rights Impact Assessment: Facebook in Myanmar."

[328] OECD Due Diligence Guidance (previously cited), p. 70.

9. META AND THE RIGHT TO REMEDY FOR THE ROHINGYA

> **“Facebook must pay. If they do not, we will go to every court in the world. We will never give up in our struggle.”**
>
> Showkutara, 22 year-old Rohingya survivor and refugee[329]

As outlined in Chapter 8, above, Meta contributed to serious adverse human rights impacts suffered by the Rohingya in the context of the 2017 atrocities in Northern Rakhine State. As a result, the company has a responsibility to provide effective remediation to affected Rohingya communities for the rights abuses the communities have suffered. This chapter surveys the efforts of Rohingya communities to seek a remedy from Meta to date.

9.1 IN PURSUIT OF JUSTICE: ROHINGYA COMMUNITIES DEMAND AN EFFECTIVE REMEDY

For the Rohingya survivors of the 2017 atrocities, the great majority of whom still live under conditions of extreme deprivation in refugee camps in Cox’s Bazar, the pursuit of justice and reparations is both a matter of principle and one of urgent material need. According to the UN Office for the Coordination for Humanitarian Affairs (UNOCHA), Rohingya refugees in Cox’s Bazar continue to face a massive funding shortfall in respect of their humanitarian needs, entailing huge impacts on their livelihoods and human rights. Of a total budget requirement of US$881million, only $117.5million, or 13.3%, has been actually funded in 2022.[330]

Since 2019, a range of complaints have been brought forward which seek justice and reparations for the 2017 atrocities suffered by the Rohingya. In November 2019, a case was filed by The Gambia against Myanmar at the International Court of Justice (ICJ), alleging violations of the Convention on the Prevention and Punishment of the Crime of Genocide (Genocide Convention).[331] In November 2019, the International

[329] Amnesty International interview by video call with Showkutara, 06 April 2022.

[330] OCHA Financial Tracking Service, ‘Bangladesh: Rohingya Refugee Crisis Joint Response Plan 2022’, fts.unocha.org/appeals/1082/summary

[331] Human Rights Watch, “Developments in Gambia’s Case Against Myanmar at the International Court of Justice”, 14 February 2022, hrw.org/news/2022/02/14/developments-gambias-case-against-myanmar-international-court-justice#whatisthe

Criminal Court (ICC)'s judges also authorized the ICC Prosecutor to open an investigation into crimes against humanity in relation to the 2017 "clearance operations".[332] Also in November 2019, a group of Rohingya survivors filed an international criminal case in Argentina under the principle of universal jurisdiction, alleging that genocide and crimes against humanity were perpetrated between 2012 and 2018 and accusing named military officials, including Senior-General Min Aung Hlaing.[333]

Alongside these criminal cases, Meta is the subject of at least three active complaints being led by Rohingya communities. Parallel civil legal proceedings were filed against Meta in December 2021 in both the United Kingdom and the United States, with the claimants seeking USD $150 billion in damages from the company.[334] In its letter of notice to Meta, lawyers representing the claimants in the UK litigation specifically allege, inter alia, that Meta "used algorithms that amplified hate speech against the Rohingya people on the Facebook platform".[335]

In addition to these court cases, groups of Rohingya youth activists located in Cox's Bazar have organized amongst themselves in pursuit of a remedy from Meta in the form of funding for educational programmes in the refugee camps.

The communities' pursuit of a remedy from Meta began on 29 July 2020, when seven camp-based Rohingya youth groups wrote to Meta's Director of Human Rights alleging that "Facebook had a large role to play in [the] violence" suffered by their community.[336] The letter sought an update on actions which Meta promised to take following the publication of its HRIA for Myanmar, and made an initial request for Meta to engage with camp-based refugees regarding the possibility of remediation. Meta organized a call with the groups in response, and during the phone call, the groups requested that Meta provide remediation. They followed up with a proposal for Meta to fund a USD $1million education project.[337] The proposal supported a pitch by a Bangladeshi academic institution to train a large number of Rohingya teachers on instructing children using the Myanmar curriculum and international materials. [338]

Educational opportunities have long been denied to the Rohingya. Under conditions of apartheid in Rakhine State, Rohingya students have been prevented from attending university since 2012.[339] The campaign of ethnic cleansing by the Myanmar military in 2017-2018, in which the authorities uprooted people by deliberately destroying their homes and belongings and forced them into exile, resulted in a host of human rights violations of economic and social rights, including the right to education, particularly for children.[340] Over half of all the refugees registered in the refugee camps in Bangladesh – among a total of approximately one million people – are children. They have been deprived of access to education in an accredited curriculum since they sought refuge in Bangladesh in 2017.[341] Since December 2021, the community's educational prospects faced further setbacks as the Bangladeshi authorities shut down or dismantled about 30 community-led schools.[342]

[332] Human Rights Watch, "Developments in Gambia's Case Against Myanmar at the International Court of Justice", 14 February 2022, hrw.org/news/2022/02/14/developments-gambias-case-against-myanmar-international-court-justice#whatisthe

[333] Burmese Rohingya Organization UK, Submission to the Federal Criminal Court of Argentina, available at: burmacampaign.org.uk/media/Complaint-File.pdf

[334] Dan Milmo, "Rohingya sue Facebook for £150bn over Myanmar genocide", 6 December 2021, theguardian.com/technology/2021/dec/06/rohingya-sue-facebook-myanmar-genocide-us-uk-legal-action-social-media-violence

[335] McCue Jury and Partners, Letter of Notice, 6 December 2021, on file with Amnesty International.

[336] Victim Advocates International, "Rohingya victim groups ask Facebook to provide them support, following its role in the violence against them", 10 July 2020, victimadvocatesinternational.org/letter-to-facebook-human-rights-head-miranda-sissons-from-rohingya-refugee-groups

[337] Letter to Facebook, 'Supporting higher education for Rohingya refugee youth', 20 November 2020, on file with Amnesty International.

[338] Letter to Facebook, 'Supporting higher education for Rohingya refugee youth', 20 November 2020, on file with Amnesty International.

[339] Amnesty International, "*Caged without a roof*", p. 70. It has been reported that Sittwe university began accepting Rohingya students again in May 2022, but it remains unclear how they can technically enrol without citizenship cards, and the freedom of movement required to physically attend the university. See: DVB English, "Sittwe University To Admit Rohingya Students After Ten Year Ban", 12 May 2022, english.dvb.no/sittwe-university-to-admit-rohingya-students-after-ten-year-ban

[340] Amnesty International, "*We will Destroy Everything*", p.137.

[341] Amnesty International and others, "Joint statement: Bangladesh: Restore and strengthen capacity of community-led schools in Rohingya camps", 28 April 2022, amnesty.org/en/latest/news/2022/04/bangladesh-restore-and-strengthen-capacity-of-community-led-schools-in-rohingya-camps

[342] Amnesty International and others, "Joint statement: Bangladesh: Restore and strengthen capacity of community-led schools in Rohingya camps", 28 April 2022, amnesty.org/en/latest/news/2022/04/bangladesh-restore-and-strengthen-capacity-of-community-led-schools-in-rohingya-camps/

According to the UN, Rohingya educational needs in 2022 total US$ 70.5million. Of this amount, a meagre 1.6% has been actually funded.[343] Mohamed Junaid, a 23-year-old Rohingya refugee, lamented the state of educational provision for Rohingya in Cox's Bazar:

> **"Though there were many restrictions in Myanmar, we could still do school until matriculation at least. But in the camps our children cannot do anything. We are wasting our lives under tarpaulin."[344]**

Mohamed, a Rohingya refugee and a volunteer teacher of mathematics and chemistry, explained these challenges to Amnesty International:

> **"It has been almost five years we have been living in the refugee camps, but we still have no opportunity to study or go to school. Our next generation is being lost day by day. At least in Myanmar we could go from Grade One to [high school] matriculation, but because of this hate speech we had to flee to Bangladesh, so we cannot see any future for our next generation. Facebook can help us to build our future to give us some opportunity."[345]**

Showkutara, 22, works with a community-based organization that conducts educational projects for young girls in the refugee camps. She explained to Amnesty International: "We are asking for help with education because, if you want to destroy a community, you just need to restrict education. And if we want to improve our lives, education is the main tool we can use to do that." Sharing her hopes for the young girls she works with, Showkutara added, "My hope for the young girls' future is for them have a chance to study... We hope they can someday go to university and become doctors, nurses, whatever they want to be".[346]

9.2 META'S REFUSAL TO REMEDIATE THE ROHINGYA

On 10 February 2021, Meta responded to the Rohingya community's request with a rejection, stating:

> **"Unfortunately, after discussing with our teams, this specific proposal is not something that we're able to support. As I think we noted in our call, Facebook doesn't directly engage in philanthropic activities.**
>
> **We do partner for [*sic.*] on some topics, but the proposal doesn't fit within the scope of the programme we typically partner on. These generally have a more direct link to our product, internet literacy, or digital empowerment."[347]**

Meta's characterisation of the Rohingya community's request as "philanthropic activities" conveys a deeply flawed understanding of the company's human rights responsibilities. The Rohingya communities have not requested charity; they are pursuing Meta to demand that the company fulfils its responsibility to remediate the severe human rights harms they have suffered and which the company contributed to. Meta's active role in the dissemination of advocacy of hatred against the Rohingya in 2017, as well as other dehumanizing and discriminatory anti-Rohingya content, meant the company contributed to a host of human rights violations by the Myanmar military, which included the right to education after people were forced to leave their homes and country.

Sawyeddollah, one of the Rohingya youth involved in the request, told Amnesty International that the rejection was a bitter blow for the community: "Their response made me really sad. They refused to acknowledge what they did to us. I know that Facebook already acknowledged they did contribute to the situation in Myanmar, but they are not accepting it when it comes to remedy."[348]

Later, in November 2020, the Rohingya youth groups again followed up with Meta to request a remedy. In a joint letter, they stated:

[343] OCHA Financial Tracking Service, 'Bangladesh: Rohingya Refugee Crisis Joint Response Plan 2022', fts.unocha.org/appeals/1082/summary

[344] Interview by video call with Mohamed Junaid, 18 April 2022.

[345] Amnesty International interview by video call with Mohamed, 11 April 2022.

[346] Amnesty International interview by video call with Showkutara, 06 April 2022.

[347] Victim Advocates International, "Rohingya victim groups ask Facebook to provide them support, following its role in the violence against them", 10 July 2020, victimadvocatesinternational.org/letter-to-facebook-human-rights-head-miranda-sissons-from-rohingya-refugee-groups

[348] Amnesty International interview with Sawyeddollah, 06 April 2022.

> **"Last year, twenty-one Rohingya groups in the camp asked Facebook to fund an education programme in the camp, that would help pay for some of our educational needs. You refused us, but we want to ask again. In recent months, we have all become aware that Facebook knew that it was being used to incite violence against minorities like us, and that your systems were removing less than 1% of violent content. We believe we are owed a remedy.**
>
> **We ask you to speak to us, as leaders in the camp, about what we need. We ask that Facebook pays for scholarships for children and young people in the camp, the teachers who currently work as volunteers, and give money to camp-based organizations to conduct education programmes. We are asking for funds to support primary, secondary and high school classes in the camp, scholarships for university students, and funds for attending short courses and trainings. Through dialogue with our leaders, we can decide together how the money will be spent."[349]**

Meta reported record revenues of over US$117 billion in 2021, including record profits of $46.7 billion. a massive 43% increase from its 2020 profits.[350] The US$ 1 million requested by the Rohingya to fund their educational initiative represents 0.002% of Meta's 2021 profits.

Mohamed Showife told Amnesty International the message he would like to deliver to Mark Zuckerberg and Meta management if he had an opportunity to address them:

> **"I would say: now you see the entire world in front of you. You, as the founder of the Metaverse, you have the resources of the entire world in your hands. I would ask him: Do you think the Rohingya people are also human beings? Do you feel we have the same worth as you? Your dream is connecting the world in front of you, but do you know what are the dreams of the Rohingya?**
>
> **The Rohingya just dream of living in the same way as other people in this world – to have the right to education, to have the possibility to become engineers, a government officer – we also have the same dreams. But you, Facebook, you destroyed our dream."[351]**

Following Meta's refusal to fund the educational programmes, the Rohingya youth groups behind the request decided to initiate a complaint against the company under the OECD Guidelines via the Irish National Contact Point (NCP).[352] The complaint was transferred to the US NCP in June 2022.[353] As of August 2022, the complaint remained under consideration. In addition to seeking funding for educational programmes, the complaint also asks that Meta "[a]djust its business model through the lens of equity, human rights, and compassion".[354]

9.3 REMEDY FOR THE ROHINGYA: ADEQUATE COMPENSATION AND SYSTEMIC CHANGE

As noted in Section 4.3, companies that have contributed to adverse human rights impacts have a responsibility to adequately remediate affected communities. The appropriate type of remediation depends on the nature of the harm, and it can take a range of forms, such as apologies, restitution, rehabilitation, financial or non-financial compensation, and punitive sanctions (whether criminal or administrative, such as fines), as well as the prevention of harm through, for example, injunctions or guarantees of non-repetition.[355] In respect of Meta's responsibility to provide a remedy to the Rohingya, although it is not within Meta's power

[349] Joint letter from Rohingya camp-based groups to Facebook, 11 November 2021, on file with Amnesty International.

[350] Meta, "Meta Reports Fourth Quarter and Full Year 2021 Results", 2 February 2022, s21.q4cdn.com/399680738/files/doc_financials/2021/q4/FB-12.31.2021-Exhibit-99.1-Final.pdf

[351] Amnesty International interview by video call with Mohamed Showife, 12 April 2022.

[352] Victim Advocates International, "Press Release: The Contributions of Facebook Have Created Hell for Us:" Rohingya Youth Share Desire for Education in Refugee Camps, Ask Facebook for Reparations", 9 December 2021, victimadvocatesinternational.org/wp-content/uploads/2022/02/Release%E2%80%94Facebook.pdf

[353] OECD Watch, "Case: Rohingya refugees supported by Victim Advocates International vs. Facebook", undated, oecdwatch.org/complaint/rohingya-refugees-vs-facebook

[354] Victim Advocates International, "Rohingya Remediation", undated, victimadvocatesinternational.org/rohingyaremediation

[355] UN Guiding Principles Interpretative Guide, p. 7.

to provide restitution directly, Meta can provide remediation in the form of compensation, rehabilitation, apologies, and guarantees of non-repetition.

Compensation is an important form of remediation where damage can be economically assessed, and in such cases, monetary compensation should be provided. The harm that can be compensated includes: "physical or mental harm" and "lost opportunities, including employment, education and social benefits."[356] The educational programmes requested by the communities in Cox's Bazar and the civil litigation claims in the US and the UK all fall into this category of remedy.

Amnesty International has not sought to quantify the losses suffered by the Rohingya in financial terms; however, both forms of remedy – financial settlements and the funding for educational programming – are valid forms of remediation under international human rights law and standards, and both could be provided by Meta in concert. Given that they have been initiated by different victims' groups, and focus on different forms of compensation, these processes should not be viewed as mutually exclusive by the competent authorities; neither should the provision of any one form of remedy be interpreted as fully satisfying Meta's responsibility to remediate the Rohingya.

Meta could additionally support remediation in the form of **rehabilitation**, which would include any medical or psychological care needed by the victims, in addition to support from legal and social services.[357] Many Rohingya refugees suffer enduring trauma based on their experiences, and provision of social services in the camps remains extremely limited. As previously noted, 86.7% of the humanitarian needs of Rohingya refugees remain unmet as of 2022.[358]

Meta could also support remediation in the form of a public **apology**, including acknowledgement of the facts and acceptance of responsibility, which could be accompanied by verification of the facts and full and public disclosure of the truth.[359] This could potentially encompass Meta's full and voluntary cooperation with the ongoing cases pursuing claims of genocide and crimes against humanity before the ICC and ICJ, encompassing the full disclosure of all relevant evidence in relation to the company's algorithmic amplification of harmful content.[360]

Lastly, **guarantees of non-repetition** are an equally important form of remediation that are intended to prevent abuses from happening again. The prevention of further abuses can be achieved through a number of measures, including both regulatory and accountability measures to be taken by states, and actions to be taken by companies - any or all of which will contribute to non-repetition in the future. [361] It could include an internal investigation into the company's specific failings in respect of the 2017 atrocities in Myanmar, including its repeated failure to act on warnings from civil society. Rohingya groups have explicitly requested a change to Meta's business model as a guarantee of non-repetition to be contemplated as part of their remedy. In Meta's case, this aspect of its responsibility to provide an effective remedy is of critical importance not only for the Rohingya – but for the human rights of at-risk communities across the world. As a global company which operates in high-risk and conflict-affected settings in every region of the world, there is a grave and present risk that Meta's operations could fuel advocacy of hatred, violence, and even genocide, of ethnic and religious minorities in many other parts of the world.

The alarm has already been raised in a number of contexts. Whistle-blower Frances Haugen has repeatedly warned that Meta is repeating its failures in Myanmar in other countries – notably including Ethiopia. Haugen specifically highlighted the Facebook platform's content-shaping algorithms as the key driver of these risks

[356] Principle 20, UN Basic Principles and Guidelines on the Right to a Remedy and Reparation for Victims of Gross Violations of International Human Rights Law and Serious Violations of International Humanitarian Law, UN Doc A/RES/60/147, 21 March 2006.

[357] Principle 21, UN Basic Principles and Guidelines on the Right to a Remedy and Reparation for Victims of Gross Violations of International Human Rights Law and Serious Violations of International Humanitarian Law, UN Doc A/RES/60/147, 21 March 2006.

[358] OCHA Financial Tracking Service, 'Bangladesh: Rohingya Refugee Crisis Joint Response Plan 2022', fts.unocha.org/appeals/1082/summary

[359] Principle 22, UN Basic Principles and Guidelines on the Right to a Remedy and Reparation for Victims of Gross Violations of International Human Rights Law and Serious Violations of International Humanitarian Law, UN Doc A/RES/60/147, 21 March 2006.

[360] Facebook has previously been accused of withholding evidence from these international tribunals. See: Robert Burnson, 'Facebook's Stance on Myanmar Genocide Records Assailed by Gambia', *Bloomberg*, 28 October 2021, bloomberg.com/news/articles/2021-10-28/facebook-s-stance-on-myanmar-genocide-records-assailed-by-gambia

[361] Injustice Incorporated: Corporate abuses and the human rights to remedy, Amnesty International, 2014, POL 30/001/2014, p. 18.

and harms.[362] In India[363] and Sri Lanka,[364] too, the spread of anti-Muslim hate and violence has been linked to Meta's content-shaping algorithms.

In order for Meta's remedy to be truly effective, the root cause of its contribution to these human rights harms – Meta's surveillance-based business model and its related use of engagement-centric algorithms - must be addressed as a matter of urgency. This reality entails urgent action by both states and technology companies alike.

A central component of the state duty to protect human rights is the obligation to enact and enforce laws and regulations which prevent and punish corporate human rights abuses. States must enact regulations which rein in surveillance-based business models throughout the technology sector, including through effective regulations regarding the use of social media ranking and recommendation algorithms.[365] These algorithms should be subject to meaningful and rigorous state oversight which have human rights law and standards at their heart. In order to tackle the root cause of these issues, such regulations must be accompanied by moves to ban tech companies' invasive tracking practices and surveillance-based advertising.[366]

While Meta has made some progress in improving its content moderation for Facebook in Myanmar, it is clearer than ever that these improvements are incapable of effectively and adequately preventing and mitigating the harms associated with Meta's content-shaping algorithms, and the voracious pursuit of personal data which is hardwired into its business model. Meta should nonetheless seek to ensure the adequate resourcing of its content moderation operations in Myanmar and in all countries and languages in which it operates.

[362] Emmanuel Akinwotu, Facebook's role in Myanmar and Ethiopia under new scrutiny, 7 October 2021, *The Guardian*, theguardian.com/technology/2021/oct/07/facebooks-role-in-myanmar-and-ethiopia-under-new-scrutiny

[363] See: Billy Perrigo, "Facebook Was Used to Incite Violence in Myanmar. A New Report on Hate Speech Shows It Hasn't Learned Enough Since Then", 29 October 2019, time.com/5712366/facebook-hate-speech-violence ;

[364] Article One, "Assessing the human rights impact of the Facebook platform in Sri Lanka", 2018, about.fb.com/wp-content/uploads/2020/05/Sri-Lanka-HRIA-Executive-Summary-v82.pdf

[365] See: Amnesty International, *Surveillance Giants* (previously cited).

[366] See: Amnesty International, *Surveillance Giants* (previously cited).

10. CONCLUSION AND RECOMMENDATIONS

> **"I'm afraid that Facebook has now turned into a beast, and not what it originally intended"**
>
> Yanghee Lee, UN Special Rapporteur on the situation of human rights in Myanmar, 2018[367]

CONCLUSION

This report, based on a thorough investigation of Meta's role in serious human rights abuses perpetrated against the Rohingya in Northern Rakhine State in 2017, has firmly established that the company substantially contributed to these harms and, therefore, has a corresponding responsibility to provide an effective remedy to affected communities.

In a context where the Rohingya had faced decades of systematic oppression, violence, and discrimination, Meta's dereliction of its responsibility to conduct appropriate human rights due diligence helped to create an enabling environment for an outbreak of mass violence against the Rohingya. The company's reality-warping algorithms actively amplified and promoted harmful content, including content which incited violence and discrimination against the Rohingya, and delivered this content directly to those most likely to act up such incitement. At the same time, the company consistently failed to act upon the repeated warnings it received from civil society stating that it risked contributing to mass violence in Myanmar, and instead allowed content targeting the Rohingya to run rampant on its platform.

Meta's reckless business practices in Myanmar lit a match amidst a tinderbox of simmering ethnic tensions, and it ultimately substantially contributed to the litany of severe human rights abuses suffered by the Rohingya. For the Rohingya, one of the most historically marginalized and oppressed minority groups in the world, Facebook indeed became a 'beast'.

Meta's response to these atrocities has only added insult to injury. The company's presentation of itself a passive and neutral platform which was 'misused' by other actors is a clear attempt to deflect from the company's active contribution to spreading and amplifying harmful content including advocacy of hatred and violence. The company's attempt to characterize Rohingya communities' efforts to secure a remedy from the company as requests for "philanthropy" is equally problematic and represents another attempt to evade responsibility. Although Meta has improved some aspects of its content moderation and community engagement in Myanmar, these reforms fail to provide guarantees of non-repetition through addressing the

[367] BBC, "UN: Facebook has turned into a beast in Myanmar", 13 March 2018, bbc.co.uk/news/technology-43385677

root cause of Meta's threat to human rights - the company's destructive business model. As this report has shown, content-based solutions will never be sufficient to prevent and mitigate algorithmic harms.

The findings of this research are not only relevant to Rohingya survivors; they should sound the alarm that Meta risks contributing to serious human rights abuses again. Across the world, Meta's content-shaping algorithms risk causing further harms in societies across the world by fanning the flames of hatred, violence and discrimination – and disproportionally impacting the most marginalized and oppressed communities, particularly in Global South contexts.

Urgent, wide-ranging reforms to Meta's business practices are needed to ensure that Meta's history with the Rohingya does not repeat itself elsewhere. It is abundantly clear, however, that Meta will not solve these problems of its own accord. The root cause of Meta's horrendous human rights impacts is hard-wired into the company's business model – a business model that is shared by other Big Tech companies. And to date, Big Tech has proven itself incapable of addressing these issues in the absence of effective regulation.

The era of unregulated Big Tech has enabled the corporate capture of our social infrastructures and information ecosystems, with grave consequences for human rights throughout the world. It is therefore essential that states fulfil their obligation to protect human rights by introducing and enforcing effective legislation to rein in surveillance-based business models and associated business practices across the technology sector.

RECOMMENDATIONS

TO META

Remedy

- Work with survivors and the civil society organizations supporting them to provide an effective remedy to affected Rohingya communities, including by:
    - Calculating and providing appropriate compensation for the Rohingya based on an assessment of physical and mental harm, lost opportunities, including employment, education and social benefits, material damages and loss of earnings, including loss of earning potential, and moral damage.
    - Supporting rehabilitation by providing for legal, medical, and psychological care needed by the victims.
    - Publicly acknowledging the full extent of Meta's contribution to human rights harms, issuing a direct apology to victims, and committing to change Meta's business model and providing remediation to similarly affected communities in other contexts.
- Cooperate fully with the OECD NCP process in the United States, and any other processes that may arise from this complaint, and fully fund the education programming requested by Rohingya communities who are parties to the complaint.

Human rights due diligence

- Undertake a comprehensive review and overhaul of human rights due diligence at Meta, including by mainstreaming human rights considerations throughout all Meta platforms' operations, especially in relation to the development and deployment of its algorithmic systems.
- Ensure that human rights due diligence policies and processes address the systemic and widespread human rights impacts of Meta's business model as a whole, and be transparent about how risks and impacts are identified and addressed.
- Elaborate professional standards for AI engineers, translating human rights responsibilities into guidance for technical design and operation choices for algorithms and other products and services.
- Ensure that human rights impact assessments are conducted in relation to the design and deployment of new AI systems, including the deployment of existing systems in new global

markets, to include meaningful public consultations and engagement prior to the finalization or roll-out of a product or service with civil society, human rights defenders, and representatives of marginalized or underrepresented communities.

- Undertake constant, ongoing and proactive human rights due diligence throughout the lifecycle of algorithmic technologies, including after the roll-out and implementation of new systems and design features, so that risks and abuses can be identified during the development stage but also after such technologies have been launched.

Business model and algorithms

- Cease the collection of invasive personal data which undermines the right to privacy and threatens a range of human rights.
- End the practice of using tracking-based advertising and embrace less harmful alternative business models, such as contextual advertising.
- To protect people's privacy and to give them real choice and control, a profiling-free social media ecosystem should not be an option but the norm. Therefore, content-shaping algorithms used by online platforms should not be based on profiling by default and must require an opt-in instead of an opt-out, with the consent for opting in being freely given, specific, informed, and unambiguous.
- Introduce 'friction' measures as a norm – not an emergency response, incorporating measures which studies have proven to be effective at improving 'integrity' outcomes, e.g., limits on resharing, message forwarding, and groups sizes.
- Radically improve transparency in relation to the use of content-shaping and content moderation algorithms, ensuring that their mechanics are publicly available in clearly understandable terms.
- Enable independent researchers to access and review algorithmic systems.
- Refrain from retiring Crowdtangle and widen access to the tool for civil society organizations, academics and journalists.

Global South

- Ensure appropriate investment in local-language resourcing throughout the world, with a particular emphasis on resolving existing inequalities that disproportionately impact Global South countries.
- Ensure equality and consistency between jurisdictions in respects of the resourcing of content moderation, policy, and human rights teams globally.

TO META'S 'HOME' STATES INCLUDING USA AND IRELAND, AND REGIONAL BODIES SUCH AS THE EU

- Ban targeted advertising on the basis of invasive tracking practices, such as cross-site tracking and tracking based on sensitive data or other personal data.
- Introduce obligations for platform companies to ensure they address systemic risks to human rights stemming from the functioning and use made of their services.
- Legally require companies, including social media companies, to conduct human rights due diligence on their business operations, products and services, as well as their business relationships and report publicly on their due diligence policies and practices in accordance with international standards.
- Regulate technology companies to ensure that content-shaping algorithms used by online platforms are not based on profiling by default and must require an opt-in instead of an opt-out, with the consent for opting in being freely given, specific, informed and unambiguous.
- Ensure adequate investment in independent oversight, monitoring, and enforcement of regulations governing the technology sector.

AMNESTY INTERNATIONAL IS A GLOBAL MOVEMENT FOR HUMAN RIGHTS. WHEN INJUSTICE HAPPENS TO ONE PERSON, IT MATTERS TO US ALL.

CONTACT US

✉ info@amnesty.org

☎ +44 (0)20 7413 5500

JOIN THE CONVERSATION

f www.facebook.com/AmnestyGlobal

@Amnesty

THE SOCIAL ATROCITY

META AND THE RIGHT TO REMEDY FOR THE ROHINGYA

Beginning in August 2017, the Myanmar security forces undertook a brutal campaign of ethnic cleansing against Rohingya Muslims in Myanmar's Northern Rakhine State. A UN investigation found that the role of Facebook in the violence was "significant".

This report is based on an in-depth investigation into Meta (formerly Facebook)'s role in the serious human rights violations perpetrated against the Rohingya. It reveals that in the months and years leading up to the 2017 atrocities, the Facebook platform became an echo chamber of virulent anti-Rohingya content in Myanmar. Meta's algorithms proactively amplified and promoted content which incited violence, hatred, and discrimination against the Rohingya – pouring fuel on the fire of long-standing discrimination and substantially increasing the risk of an outbreak of mass violence.

Despite its partial acknowledgement that it played a role in the 2017 violence against the Rohingya, Meta has to date failed to provide an effective remedy to affected Rohingya communities. However, Amnesty International's systematic legal analysis of Meta's role in the atrocities perpetrated against the Rohingya leaves little room for doubt: Meta substantially contributed to adverse human rights impacts suffered by the Rohingya and has a responsibility to provide survivors with an effective remedy.

INDEX: ASA 16/5933/2022
AUGUST 2022
LANGUAGE: ENGLISH

amnesty.org

AMNESTY
INTERNATIONAL