

Flygtningenævnets baggrundsmateriale

Bilagsnr.:	559
Land:	Etiopien
Kilde:	Amnesty International
Titel:	"A death sentence for my farther" – Meta's contribution to human rights abuses in northern Ethiopia
Udgivet:	31. oktober 2023
Optaget på baggrundsmaterialet:	27. februar 2024



“A DEATH SENTENCE FOR MY FATHER”

META'S CONTRIBUTION TO HUMAN RIGHTS ABUSES
IN NORTHERN ETHIOPIA

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.

© Amnesty International 2023

Except where otherwise noted, content in this document is licensed under a Creative Commons (attribution, non-commercial, no derivatives, international 4.0) licence.

<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>

For more information please visit the permissions page on our website: www.amnesty.org

Where material is attributed to a copyright owner other than Amnesty International this material is not subject to the Creative Commons licence.

First published in 2023

by Amnesty International Ltd

Peter Benenson House, 1 Easton Street

London WC1X 0DW, UK

Index: AFR 25/7292/2023

Original language: English

amnesty.org



Cover image: Illustration by Alina Pechenckina

AMNESTY
INTERNATIONAL



CONTENTS

1. EXECUTIVE SUMMARY	5
2. METHODOLOGY	10
3. BACKGROUND	12
3.1 ETHIOPIA'S CHANGING POLITICAL LANDSCAPE	12
3.2 CONFLICT IN WESTERN TIGRAY: REVERBERATIONS THROUGHOUT ETHIOPIA	13
3.3 THE FACEBOOK PLATFORM'S IMPORTANCE IN ETHIOPIA'S INFORMATION ECOSYSTEM	16
4. LEGAL FRAMEWORK	18
4.1 BUSINESS AND HUMAN RIGHTS STANDARDS	18
4.2 HUMAN RIGHTS DUE DILIGENCE AND SOCIAL MEDIA'S ALGORITHMIC TECHNOLOGIES	19
4.3 THE CORPORATE RESPONSIBILITY TO PROVIDE REMEDY	20
4.4 THE PROHIBITION OF ADVOCACY OF HATRED UNDER INTERNATIONAL HUMAN RIGHTS LAW	21
5. FANNING THE FLAMES: META'S ROLE IN STOKING ETHNIC TENSIONS IN ETHIOPIA	23
5.1 THE ROLE OF FACEBOOK IN OFFLINE VIOLENCE IN ETHIOPIA BEFORE THE ARMED CONFLICT	23
5.2 THE ROLE OF PUBLIC FIGURES AND GOVERNMENT ALLIES IN SPREADING CONTENT TARGETING THE TIGRAYAN COMMUNITY	26
5.3 ANALYSIS: HOW ADVOCACY OF HATRED TARGETING THE TIGRAYAN COMMUNITY LINKED TO OFFLINE VIOLENCE	27
5.4 CONTENT MODERATION AND SPEECH TARGETING TIGRAYANS ON THE FACEBOOK PLATFORM	29
5.5 CONTENT MODERATION: THE HUMAN COST	32
6. ONLINE RHETORIC, OFFLINE HARM	34
6.1 THE CASE OF MEAREG AMARE	34
6.2 THE CASE OF FISSEHA TEKLE	37
6.3 THE CASE OF ITSAY ASSEFA	38
6.4 THE CASE OF LUCY KASSA	40

7. MORE THAN AN ABERRATION: META'S SYSTEMIC FAILURES IN ETHIOPIA	42
7.1 META'S POST-2017 BUSINESS MODEL: ENGAGEMENT BY ANOTHER NAME	42
7.2 RESHARES AND VIRALITY	44
7.3 MSI PRIORITIZED OVER "INTEGRITY MEASURES"	45
7.4 UNHEEDED WARNINGS	46
7.5 FAILURE TO ADEQUATELY MITIGATE RISKS	50
7.6 META'S KNOWLEDGE OF THE RISKS OF ALGORITHMIC AMPLIFICATION	51
7.7 META'S CONTRIBUTION TO HUMAN RIGHTS HARMS SUFFERED BY THE TIGRAYAN COMMUNITY	53
8. MAKE IT RIGHT: REMEDY AND PREVENTION OF FUTURE HARM	55
8.1 JUSTICE AND ACCOUNTABILITY: CIVIL LITIGATION IN KENYA	55
8.2 META'S RESPONSIBILITY TO PROVIDE REMEDY	56
8.3 RISK OF FUTURE HARM IN ETHIOPIA	57
8.3.1 ETHIOPIA'S TELECOMS EXPANSION	57
8.3.2 RISK OF FUTURE ATROCITIES IN ETHIOPIA	58
9. CONCLUSION AND RECOMMENDATIONS	60
CONCLUSION	60
RECOMMENDATIONS	62
TO META	62
TO META'S HOME STATES, INCLUDING USA AND IRELAND AND REGIONAL BODIES SUCH AS THE EU	63
APPENDIX 1	64
META RESPONSE TO AMNESTY INTERNATIONAL	64
APPENDIX 2	72
SAMA RESPONSE TO AMNESTY INTERNATIONAL	72

1. EXECUTIVE SUMMARY

“I knew it would be a death sentence for my father”, Abrham Meareg told Amnesty International, describing his reaction to seeing posts on the Facebook platform targeting his father, Professor Meareg Amare, due to his Tigrayan identity. The posts contained Professor Meareg Amare’s name, photo, place of work and made allegations about him being affiliated to the Tigray People’s Liberation Front (TPLF).

Sadly, Abrham’s instinct was correct. A few weeks after the posts were circulated – and despite Abrham’s repeated attempts to have them taken down through Meta Platforms Inc. (Meta)’s reporting function – Professor Meareg Amare was killed outside of his home on 3 November 2021.

On 4 November 2020, armed conflict broke out in Ethiopia’s Tigray region between forces aligned with Ethiopia’s federal government and forces affiliated with Tigray’s regional government, and later spread to the neighbouring Amhara and Afar regions. The conflict was characterized by extreme brutality against civilians enacted by all parties to the conflict.

A week after the conflict broke out, Amnesty International published a report showing that scores of civilians had been killed in Mai-Kadra town, in Western Tigray. Amnesty International and Human Rights Watch also found that war crimes and crimes against humanity including the crime of ethnic cleansing were perpetrated against Tigrayan civilians in the Western Tigray region by state officials and security forces from the neighbouring Amhara region, with the acquiescence and possible participation of Ethiopian federal forces.

The violence has had a devastating impact on civilians living in areas where active hostilities occurred in northern Ethiopia, including in the Tigray region. It is estimated that the conflict has resulted in the deaths of up to 600,000 civilians. Millions of Ethiopians have been internally displaced, and an estimated 70,000 people who fled are now living in refugee camps in eastern Sudan.

During the conflict, Facebook (owned by Meta) in Ethiopia became awash with content inciting violence and advocating hatred. Content targeting the Tigrayan community was particularly pronounced, with the Prime Minister of Ethiopia, Abiy Ahmed, pro-government activists, as well as government-aligned news pages posting content advocating hate that incited violence and discrimination against the Tigrayan community, planting the idea that even Tigrayan civilians presented a threat to Ethiopia’s stability and security. The mass dissemination of these messages, which incited violence and discrimination, as well as other dehumanizing and discriminatory content targeting the Tigrayan community, poured fuel on the fire of an already polarized society with significant ethnic tensions.

The rapid dissemination of these messages was made possible due to Meta’s surveillance-based business model, which is predicated on harvesting, analysing and profiting from people’s data, and privileges ‘engagement’ at all costs. This business model relies on incentivizing people to stay on the platform for as long as possible in order to collect ever-more data on them for the purpose of targeted advertising. In furtherance of this aim, Meta’s content-shaping algorithms are tuned to maximize engagement, and to boost content that is often inflammatory, harmful and divisive, as this is what tends to garner the most attention from users. In the context of the northern Ethiopia conflict, these algorithms fuelled devastating human rights impacts, amplifying content targeting the Tigrayan community across Facebook, Ethiopia’s most popular social media platform – including content which advocated hatred and incited violence, hostility and discrimination.

Ethiopian journalists and researchers also flagged a tide of hate directed at them on Facebook, simply for doing their job of documenting and reporting on the armed conflict. Ethiopian journalist Lucy Kassa told Amnesty International that she fled to Nairobi after being targeted with posts inciting violence against her after reporting on human rights violations perpetrated by government forces in Tigray. However, even in

Kenya, she was not safe from the harmful impacts of Meta's algorithms – after reporting on a possible incendiary weapons attack in Tigray, she was once again targeted, this time by a government account with hundreds of thousands of followers, which called for her to be arrested and posted her photo. The post was liked and commented on thousands of times. Lucy told Amnesty International: "the government and influential activists were calling for violence against me, saying that I should be stoned, that the government should "do something" about me. So, it wasn't safe for me in Nairobi."

Lucy was forced to flee once again and told Amnesty International that it took a long time for her to be able to confront the impact that the online violence had on her mental health.

In April 2020, before the outbreak of armed conflict in northern Ethiopia, the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression noted the importance of the Facebook platform to Ethiopia's information ecosystem and advised that "[Meta should]...deliver localized support to its burgeoning user base to ensure that its platform contributes to people's expressions, rather than becoming a tool for the spread of hate and disinformation."

In October 2023, the UN-appointed International Commission of Human Rights Experts on Ethiopia (ICHREE) concluded in its final report that "[t]he prevalence of hate speech in Ethiopia, in particular online, stoked community tensions and created a climate in which individuals and groups became targets of incitement and calls to violence." The ICHREE recommended that "[t]he full extent to which the hate speech, as well as mis- and disinformation online, has contributed to or exacerbated discrimination and violence in Ethiopia – both during the conflict and in the present day – merits further independent investigation, so that appropriate lessons can be drawn to prevent future recurrence." This report provides precisely such an independent investigation into Meta's role in the violence.

Internal documents revealed as part of the Facebook Papers, a series of documents released by a former Meta employee in 2021, demonstrate that during the northern Ethiopia conflict Meta had knowledge that the company risked contributing to human rights harms in the country but was failing to adequately mitigate these risks, despite the country being designated a priority by the company. A 2021 internal document designated Ethiopia as a country at "dire" risk of violence and in its assessment of the company's response to violent and inciting content, it ranked its own capacity in the country as a zero out of three. In another document, a Meta staff member acknowledged that the company lacked "human review capacity" for Ethiopia in the run up to the country's 2021 elections. Over the course of the conflict, civil society groups and digital rights experts sounded the alarm over the Facebook platform's risk of contributing to human rights violations and abuses against the Tigrayan community in Ethiopia. This report examines their concerns and allegations and provides an in-depth human rights analysis of the role played by Meta in the egregious human rights violations and abuses perpetrated against the Tigrayan community from 2020 to 2022, and Meta's responsibility to prevent or mitigate future harms in Ethiopia, which despite the cessation of hostilities between the Ethiopian government and the TPLF, remains a conflict-affected setting.

Gelila, who works at an Ethiopian civil society organization and was part of Meta's 'Trusted Partner' programme during the northern Ethiopia conflict – an initiative which aims to provide selected civil society groups with a designated channel to alert Meta to harmful content on its platforms – emphasized the high possibility that without immediate reforms to its operations, the Facebook platform will contribute to further human rights abuses:

"As someone who has been in Ethiopia for a long time, I can say that Facebook is making communities more vulnerable to conflict with each other."

This report outlines how Meta – through its content-shaping algorithms and data-hungry business model – contributed to the serious human rights violations and abuses perpetrated against the Tigrayan community. It reveals that Meta ignored warnings from civil society actors and did not take adequate mitigation measures even once the conflict had broken out. These failures, combined with content-shaping algorithms tuned to promote inflammatory content to increase user engagement, led to the Facebook platform becoming rife with content that advocated hatred constituting incitement to violence, hostility and discrimination against the Tigrayan community.

Meta's contribution to human rights violations and abuses in Ethiopia came three years after the 2017 atrocities against the Rohingya in Myanmar, to which the Facebook platform contributed significantly. The fact that Meta yet again contributed to human rights harms is further evidence that its business model, based on invasive profiling and targeted advertising, fuels the spread of harmful content. In the never-ending pursuit of ever-more data, Meta's algorithmic systems prioritise the most inflammatory, divisive and harmful content, which is most likely to maximize engagement.

In a typical example of the content targeting the Tigrayan community which was burgeoning on the Facebook platform at the time, Prime Minister Abiy Ahmed posted in July 2021 calling Tigray forces “weeds” and a “cancer”. As of October 2023, the post remains visible on Facebook. The post has been shared over 8,000 times, and received thousands of comments, such as: “Let the enemy be destroyed. Ethiopia will come first”. It is highly significant that the Prime Minister himself was using the Facebook platform to post content inciting hate and discrimination against the Tigrayan community. While Meta has previously stated that it does not wish to be the arbiter of whether a politician’s speech is appropriate or not, the company has also acknowledged that content which has the potential to incite violence may pose a safety risk that outweighs the public interest.

Prior to the conflict in northern Ethiopia and as it unfolded, civil society activists and digital rights experts repeatedly pleaded with Meta to take action to address emerging risks. This report documents the repeated communications and interventions received by Meta between 2019 and 2022, including by its own Oversight Board. Despite these efforts, and its record of contributing to serious violations against the Rohingya in Myanmar, Meta failed to heed these warnings or even to adequately acknowledge them.

Furthermore, Meta’s wholly inadequate investment in its moderation of Ethiopian content prior to and during the northern Ethiopia conflict was a significant factor in the company’s failure to remove harmful content targeting the Tigrayan community from the platform. While there are 85 languages spoken in Ethiopia, Meta is only capable of moderating content in four of them. This is indicative of the company’s broader failure to adequately invest in content moderation in non-English speaking Global Majority countries. The ICHREE also found indications that the Facebook platform “was too slow to respond and suffered from inadequate staffing, insufficient language capabilities, and a lack of financial investment.”

The risk that Meta could contribute to mass violence in Ethiopia should have been clear before the outbreak of the 2020 armed conflict in northern Ethiopia. In June 2020, after the murder of the popular Oromo singer and activist Hachalu Hundesa the platform saw a surge of content advocating hatred and inciting violence, calls for “revenge,” leading to a wave of brutal mob violence. Soon after this incident Meta translated its community standards into Amharic for the first time. However, this fell far short of the measures needed to adequately mitigate the full range of risks presented by the Facebook platform.

Civil society groups also tried to reach out to Meta to flag the risk of violence that the Facebook platform presented in Ethiopia. In June 2020, four months before the outbreak of the conflict in northern Ethiopia, digital rights organizations sent a letter to Meta about the harmful content circulating on Facebook in Ethiopia, warning that such content “can lead to physical violence and other acts of hostility and discrimination against minority groups.” The letter made several recommendations to prevent harms, including ceasing the algorithmic amplification of content inciting violence, temporary changes to sharing functionalities and conducting a human rights impact assessment into the company’s operations in Ethiopia.

Gelila highlighted Meta’s slow responses and lack of respect for local knowledge:

“They are extremely slow in reacting to things. They are not sensitive to what is said – I think they have standards which are very far from what is happening on the ground. When you’re local you know what things trigger what, it might not work for their policy on whether it’s hate speech or not, they might have their own understanding in their office somewhere in the West, but on the ground, you know what is hate speech in the local context.”

Meta has continuously failed to take adequate responsibility for its impacts in Ethiopia. In a decision published on 14 February 2021, the Facebook Oversight Board recommended that Meta conduct an independent human rights due diligence assessment on how the Facebook platform in Ethiopia had been used to spread unverified rumours and hate, which heightened the risk of violence in the country. This was recommended separately to the 2021 human rights impact assessment that Meta undertook on risks ahead of the Ethiopia elections. In January 2022, Meta responded that it would “assess the feasibility” of commissioning such a human rights impact assessment in Ethiopia on the spread of unverified rumours and hate. However, as of October 2023, there is no evidence of any such assessment or indeed that work on such an assessment has begun.

In December 2017, just months after the company played a high-profile role in the ethnic cleansing perpetrated against the Rohingya people of Myanmar, Meta pivoted its News Feed algorithm around a new metric, dubbed “meaningful social interactions” (MSI). This change sought to counter declining engagement on the platform and was presented as an effort to “fix Facebook” by CEO Mark Zuckerberg. Yet this report finds that MSI is simply engagement by another name and that it presents significant risks, particularly in conflict-affected settings. The metric encourages comments and user interactions, but it is still premised on

engagement-based ranking – and internal Meta documents suggest that MSI results in algorithmic boosts for content that is inflammatory and harmful.

The Facebook Papers contain evidence suggesting that Meta’s CEO Mark Zuckerberg personally intervened to stop the application of proposed mitigation measures due to concerns around potential negative impacts of those measures on MSI, and therefore, the profitability of the Facebook platform. In a document discussing ‘soft actions’ (an umbrella term for a wide range of options that Meta can apply that stop short of content removal, for content in crisis situations such as Ethiopia), the author writes: “Mark doesn’t think we should go broad... We wouldn’t launch if there were material trade-offs with MSI”. In this document, Ethiopia is mentioned as a country where experiments with ‘soft actions’ will be temporarily launched, due to its status as an at-risk country.

Ethiopia was also specifically highlighted by Facebook whistle-blower Frances Haugen during her 2021 Congressional Testimony when she discussed Meta’s responsibility for algorithmic boosting, saying:

“[Meta] knows, they have admitted in public, that engagement-based ranking is dangerous without integrity and security systems, but then not rolled those systems out to most languages in the world... And that’s what’s causing things like ethnic violence in Ethiopia.”

All companies have a responsibility to respect human rights wherever they operate in the world and throughout their operations. 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 such as Meta, due diligence should also include addressing situations in which their business model and design decisions create or exacerbate human rights risks.

Meta failed to conduct appropriate human rights due diligence in respect of its operations in Ethiopia ahead of the 2020 northern Ethiopia conflict, and even after its own Oversight Board recommended a human rights impact assessment be conducted in the country. This analysis leaves little room for doubt that Meta has, once again, contributed to serious human rights violations and abuses in a conflict-affected setting.

On 18 July 2023, Amnesty International wrote to Meta and asked what mitigation measures the company has taken in Ethiopia prior to and during the northern Ethiopia conflict to prevent the Facebook platform from contributing to human rights abuses. Meta responded outlining their due diligence approach and stakeholder engagement in Ethiopia since 2020. Amnesty International wrote to Meta again in October 2023 regarding the allegations in this report. Meta made it clear that they disagreed with the findings of the report but stated that were unable to comment further due to pending litigation.

To this day, Meta continues to fail to adequately address risks that its operations present in Ethiopia. However, it is more urgent than ever that the company takes meaningful steps to properly mitigate potential future harms: there is currently a crisis developing in the Amhara region and simmering ethnic tensions exist across Ethiopia. This, combined with the expansion of the telecommunications network underway in the country, means that conditions are ripe for the Facebook platform to once again be used as a vector of hate, violence and discrimination.

Meta is currently facing civil litigation in Kenya for its alleged contribution to human rights abuses in Ethiopia. The litigation includes two individual petitioners, Abraham Meareg and Fisseha Tekle, and a Kenyan civil society organization, the Katiba Institute, who are seeking to stop the Facebook platform’s algorithms from recommending harmful content and to compel Meta to create a USD 1.6 billion victims’ fund to begin repairing the damage it has contributed to.

As detailed throughout this report, Meta’s failure to uphold its human rights responsibilities, as outlined in the UN Guiding Principles on Business and Human Rights (UNGPs), has contributed to devastating impacts for the Tigrayan community. This devastation is compounded by the fact that Meta has already been found to have contributed to serious human rights abuses in Myanmar and similar systemic failures have been repeated in Ethiopia. It cannot be any clearer that Meta’s business practices present a significant danger to human rights, particularly in conflict-affected settings. Urgent, wide-ranging reforms are needed to ensure that Meta does not contribute again to these harms in yet another country – including, crucially, that Meta commits to changing its surveillance-based business model to ensure that human rights harms are not repeated in the future.

Meta’s repeated failures in Ethiopia demonstrate that the company is still failing to address the source of its negative human rights impacts. The root cause of Meta’s repeated contribution to human rights violations is its surveillance-based business model, which is shared by several other Big Tech companies. These issues can only be addressed with robust and meaningful state regulation and oversight. States must fulfil their

obligation to protect human rights by introducing and enforcing legislation to effectively rein in Big Tech companies' surveillance-based business model.

Abrham Meareg, the grieving son of Professor Meareg Amare, expressed his motivation for engaging in civil litigation against Meta and underlined the urgency that the company finally learns a lesson from its operations in Ethiopia in simple terms:

“We can save so many lives.”

2. METHODOLOGY

This report is based on research conducted by Amnesty International between February and July 2023. Amnesty International conducted an analysis of the human rights implications of Meta's business operations in Ethiopia using information from the Facebook Papers, a cache of internal Meta documents which were disclosed by the whistle-blower Frances Haugen to the United States (US) Congress in October 2021. These documents were obtained from public sources, including those published by Gizmodo.¹ This analysis was informed by Amnesty International's interviews with subject matter experts and by desk research. The organization also carried out extensive desk research from open sources, including relevant international human rights standards, reports from civil society organizations, international news media, and United Nations (UN) reports.

The report builds on extensive previous investigations by Amnesty International and a joint investigation between Amnesty International and Human Rights Watch that documented crimes against humanity and ethnic cleansing committed against the Tigrayan community during the armed conflict in northern Ethiopia between 2020 and 2022.² It also builds on Amnesty International's previous research on the impact of Meta's content-shaping algorithms in Myanmar during the 2017 atrocities against the Rohingya community, and the human rights impacts of Big Tech's surveillance-based business model.³

Amnesty International conducted remote interviews with six affected individuals. Two of the affected individuals interviewed are the petitioners in a civil litigation case against Meta filed in Kenya – including an Amnesty International staff member, who is being interviewed in his personal capacity. Amnesty International is acting as an interested party in the case. The Kenyan Constitutional Court defines an interested party as a person or entity that has an identifiable stake or legal interest or duty in the proceedings before the court but is not a party to the proceedings or may not be directly involved in the litigation.⁴ It should be noted that the framework of analysis for Meta's contribution to human rights harms used in this report is grounded in international human right standards, including the UN Guiding Principles on Business and Human Rights. This report does not seek to assess Meta's responsibility for any harms under Kenyan legal standards.

The four case studies which feature in this report are illustrative and emblematic of the harms the Facebook platform contributed to in Ethiopia. The organization conducted nine further interviews with subject matter experts, including digital rights experts, journalists and content moderators.

Most interviews were conducted in English, while two were conducted in Amharic with the support of an interpreter. The information gathered from these interviews was then corroborated with local digital rights organizations, news coverage 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

¹ Dell Cameron and others, "Read the Facebook Papers for Yourself", Gizmodo, 18 April 2022, <https://gizmodo.com/facebook-papers-how-to-read-1848702919>

² Amnesty International & Human Rights Watch, *We will erase you from this land: Crimes against humanity and ethnic cleansing in Ethiopia's Western Tigray Zone* (Index: AFR 25/5444/2022), 6 April 2022, <https://www.amnesty.org/en/documents/afr25/5444/2022/en/>

³ Amnesty International, *The Social Atrocity: Meta and the right to remedy for the Rohingya* (Index: ASA 16/5933/2022), 29 September 2022, <https://www.amnesty.org/en/documents/asa16/5933/2022/en/>; Amnesty International, *Surveillance giants: How the business model of Google and Facebook threatens human rights* (Index: POL 30/1404/2019), 21 November 2019, <https://www.amnesty.org/en/documents/pol30/1404/2019/en/>

⁴ Kenya Law, *Land Case No.27 of 2021*, <http://kenyalaw.org/caselaw/cases/view/221027#:~:text=Rule%20of%20the%20Mutunga,para.14>, para. 14.

share their identities publicly. For those who chose anonymity, Amnesty International used pseudonyms and omitted all potentially identifying information from this report.

On 18 July 2023, Amnesty International wrote to Meta and asked questions regarding the company's actions in relation to its business activities in Ethiopia before and during the armed conflict in northern Ethiopia between 2020 and 2022.⁵ Meta responded outlining their due diligence approach and stakeholder engagement in Ethiopia, as well as providing information on their efforts to improve content moderation since 2020.⁶ Amnesty International wrote to Meta again on 11 October 2023 to inform the company of relevant allegations contained in this report and to give the company the opportunity to respond.⁷ Meta responded to Amnesty International on 25 October 2023, making it clear that they 'fundamentally disagree' with the report's findings, highlighting that Ethiopia is a priority country for intervention and noting that they were unable to respond more fully due to pending litigation.⁸ Their response is included in this report's Annex.

Throughout this report, the term "Meta" is used to refer to the company, Meta Platforms Inc. (formerly Facebook Inc.), including in relation to the period before the company's rebrand 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.

⁵ Amnesty International letter to Meta, 18 July 2023.

⁶ Meta letter to Amnesty International, 18 September 2023

⁷ Amnesty International letter to Meta, 11 October 2023

⁸ Meta letter to Amnesty International, 26 October 2023 – see Annex below

3. BACKGROUND

3.1 ETHIOPIA'S CHANGING POLITICAL LANDSCAPE

The Tigray People's Liberation Front (TPLF) was formed in the 1970s, mobilizing against the monarchy and later the military regime, Derg, that ruled Ethiopia from 1974 to 1991.⁹ The TPLF, alongside other political parties and rebel fronts, engaged in a series of armed revolts against the Derg.¹⁰ At the peak of the conflict and amidst a famine, between 1984 and 1985, the then government forcibly relocated hundreds of thousands of people to well-watered regions in the south and the west.¹¹ The Eritrean People's Liberation Front (EPLF) also fought against the Derg regime alongside the TPLF and other armed groups in Ethiopia.¹² After 17 years of armed resistance, these groups, operating under the umbrella of the Ethiopian People's Revolutionary Democratic Front (EPRDF), captured Addis Ababa and overthrew the Derg regime following the departure of then leader Mengistu Hailemariam to Zimbabwe, where he still resides.¹³

The EPRDF, a coalition of political parties from four regional states of the country (Amhara, Oromia, Tigray and Southern Nations and Nationalities), ruled the country from 1991 until 2019.¹⁴ The EPRDF, which was dominated by the TPLF, governed Ethiopia until it faced widespread protests, first in the Oromia region and later in the Amhara region.¹⁵ The EPRDF-led government was known for its repressive tactics, including harsh and severe suppression of opposition party members, the media, human rights defenders, as well as domestic and international non-governmental organizations.¹⁶

After waves of protests in the country since 2015 and following an internal power struggle, the EPRDF elected Abiy Ahmed from the Oromia wing of the coalition as its chairperson, leading to his confirmation by the parliament as Prime Minister of Ethiopia in 2018.¹⁷

The new government continued the reform process initiated by Abiy Ahmed's predecessor, which included the return of exiled political parties, release of detained opposition politicians, activists and journalists and the revision of repressive laws.¹⁸

⁹ The Guardian, "Rise and fall of Ethiopia's TPLF – from rebels to rulers and back", 25 November 2020, <https://www.theguardian.com/world/2020/nov/25/rise-and-fall-of-ethiopias-tplf-tigray-peoples-liberation-front>

¹⁰ Gebru Tareke, "The Ethiopian Revolution: War in the Horn of Africa", 2013, African Studies Review, Vol52 Issue 3; John Markakis, *Ethiopia: The Last Two Frontiers*, 2011.

¹¹ The New York Times, "Ethiopia moving 1.5 million people from famine", 14 December 1984, <https://www.nytimes.com/1984/12/14/world/ethiopia-moving-1.5-million-people-from-famine.html>; BBC News, "World still learning from Ethiopia famine", 29 November 2014, <https://www.bbc.co.uk/news/world-africa-30211448>

¹² Crisis Group, "Eritrea's long bitter feud with Ethiopia's Tigray region", 9 November 2022, <https://www.crisisgroup.org/africa/horn-africa/eritrea-ethiopia/eritreas-long-bitter-feud-ethiopias-tigray>

¹³ The Guardian, "Rise and fall of Ethiopia's TPLF leaders – from rebels to rulers and back", 25 November 2020, <https://www.theguardian.com/world/2020/nov/25/rise-and-fall-of-ethiopias-tplf-tigray-peoples-liberation-front>

¹⁴ The Guardian, "Rise and fall of Ethiopia's TPLF – from rebels to rulers and back" (previously cited).

¹⁵ Al Jazeera, "Ethiopia: Mass protests 'rooted in country's history'", 20 February 2018, <https://www.aljazeera.com/news/2018/2/20/ethiopia-mass-protests-rooted-in-countrys-history>

¹⁶ Amnesty International, "Ethiopia: Lack of accountability for past violations haunts the present", 28 May 2019, <https://www.amnesty.org/en/latest/press-release/2019/05/ethiopia-lack-of-accountability-for-past-violations-haunts-the-present/>; Committee to Protect Journalists, "Ethiopia archive", <https://cpj.org/africa/ethiopia/>

¹⁷ Al Jazeera, "Abiy Ahmed sworn in as Ethiopia's prime minister", 2 April 2018, <https://www.aljazeera.com/news/2018/4/2/abiy-ahmed-sworn-in-as-ethiopias-prime-minister>

¹⁸ Amnesty International & Human Rights Watch, *We will erase you from this land: Crimes against humanity and ethnic cleansing in Ethiopia's Western Tigray Zone* (previously cited).

One year after coming into power, Prime Minister Abiy Ahmed dissolved the EPRDF and formed a new national political party called Prosperity Party, promising to move away from “Ethiopia’s ethnically divisive politics.”¹⁹

One Tigrayan interviewee described this as a time of hope for some in Ethiopia, saying “I was supporting the changes, all the reforms in Ethiopia.”²⁰

However, these reforms sparked the deterioration of relations between the TPLF and the central government. The TPLF refused to join the new Prosperity Party, declaring that the merger was an “illegal move.”²¹

The TPLF opposed prosecutions of party leaders for past human rights violations.²² Federal government officials accused the TPLF of stoking unrest and supporting ethnic violence and armed insurgencies in Ethiopia.²³

After rejecting the postponement of Ethiopia’s August 2020 elections due to the Covid-19 pandemic, the TPLF regional administration unilaterally conducted a regional election in Tigray in September 2020, in which it was declared winner in all regional seats except one.²⁴ In response, the federal government declared the election null and void, ordered that communication between the federal and the Tigray regional government be cut off and froze fund transfers to the Tigray government.²⁵ Attempts to mediate between the TPLF and the federal government were unsuccessful. TPLF officials wanted the formation of a transitional government, a comprehensive national dialogue, and release of political opposition leaders, which the federal government rejected.²⁶

The political friction, combined with decades of resentment around TPLF power and influence, created a hostile environment towards the Tigrayan community. As one Tigrayan community member told Amnesty International: “The Tigrayans became a scapegoat in the country. There was blame-shifting. It was all politics.”²⁷

3.2 CONFLICT IN WESTERN TIGRAY: REVERBERATIONS THROUGHOUT ETHIOPIA

Amidst a backdrop of simmering political tensions, on 4 November 2020 Prime Minister Abiy Ahmed announced on national TV that Tigray security forces had attacked the Ethiopian National Defense Forces (ENDF) Northern Military Command at several military bases in the region, and that consequently he had launched a “law enforcement operation” against the Tigray regional paramilitary police and militia loyal to the TPLF in response.²⁸ The armed conflict that followed was characterized by extreme brutality, with Amnesty International finding that all parties committed war crimes.²⁹ These findings also included that over the course of the conflict, crimes against humanity and war crimes were committed by civilian authorities and Amhara regional security forces with the acquiescence and possible participation of Ethiopian federal forces against Western Tigray’s Tigrayan population including murder, enforced disappearances, torture, deportation and forcible transfer, rape, sexual slavery and other sexual violence, persecution, unlawful imprisonment, possible extermination and other unlawful acts.³⁰

¹⁹ BBC, “Ethiopia’s Abiy Ahmed gets a new ruling party” (previously cited).

²⁰ Amnesty International interview by video call with Abrham Meareg, 11 April 2023.

²¹ Addis Standard, “News: TPLF cites legal, procedural shortfalls to decline merger”, 21 November 2019, <https://addisstandard.com/news-tplf-cites-legal-procedural-shortfalls-to-decline-merger/>

²² Amnesty International, *I don’t know if they realized I was a person: Rape and other sexual violence in Tigray, Ethiopia* (Index: AFR 25/4569/2021), 11 August 2021, <https://www.amnesty.org/en/documents/afr25/4569/2021/en/>

²³ Amnesty International, *I don’t know if they realized I was a person: Rape and other sexual violence in Tigray, Ethiopia* (previously cited).

²⁴ Al Jazeera English, “Governing party in Ethiopia’s Tigray sweeps regional polls”, 11 September 2020,

<https://www.aljazeera.com/news/2020/9/11/governing-party-in-ethiopias-tigray-sweeps-regional-polls>

²⁵ Al Jazeera English, “Ethiopian parliament votes to cut ties with Tigray region leaders”, 7 October 2020,

<https://www.aljazeera.com/news/2020/10/7/ethiopia-parliament-votes-to-cut-ties-with-tigray-region-leaders>

²⁶ Crisis Group, *Keeping Ethiopia’s Transition on the Rails*, Report No. 283, 16 December 2019, <https://www.crisisgroup.org/africa/horn-africa/ethiopia/283-keeping-ethiopias-transition-rails>

²⁷ Amnesty International interview by video call with Abrham Meareg, 11 April 2023.

²⁸ BBC: “Ethiopia orders military response after army base seized”, 4 November 2020, <https://www.bbc.com/news/world-africa-54805088>

²⁹ Amnesty International & Human Rights Watch, *We will erase you from this land: Crimes against humanity and ethnic cleansing in Ethiopia’s Western Tigray Zone* (previously cited); Amnesty International, *Ethiopia: Summary killings, rape and looting by Tigrayan forces in Amhara* (Index: AFR 25/5218/2022), 16 February 2022, <https://www.amnesty.org/en/documents/afr25/5218/2022/en/>

³⁰ Amnesty International & Human Rights Watch, *We will erase you from this land: Crimes against humanity and ethnic cleansing in Ethiopia’s Western Tigray Zone* (previously cited).

A day after the conflict broke out, the Ethiopian government enforced an internet and telecommunications shutdown in the Tigray region.³¹ The region was cut off from the rest of the country and the wider world. Humanitarian organizations attempting to operate in the area highlighted that access and telecommunications limitations in most parts of Tigray made it difficult to assess both the full impact of the ongoing hostilities and the security of humanitarian workers in the region.³² Internet shutdowns directly violate multiple human rights, including the rights to freedom of expression and information, while also hindering other rights such as the right to freedom of peaceful assembly.³³

Western Tigray, where fighting initially began, saw the widespread pillage of crops and livestock, as well as the destruction and occupation of Tigrayan homes by Amhara security forces. Tigrayans faced mass arrest by Amhara security forces and civilian administrators and prolonged arbitrary detention in formal and informal detention sites where detainees were killed, tortured and otherwise ill-treated. Regional authorities imposed discriminatory rules that denied Tigrayans basic services and access to humanitarian aid and measures that appeared designed to suppress their rights and presence in the area.³⁴ Tigrayans also endured ethnic-based slurs that targeted their identity and were banned from speaking their language.³⁵

In several towns including Humera, Ruwassa, Adi Goshu, Adebai and Baker, the plans to remove Tigrayans from the area were a matter of public discussions and displays.³⁶ Interim authorities and security officials repeated slogans such as “Tigrayans belong east of the Tekeze river”, and “this is Amhara land”, underscoring that Tigrayans were being pushed out of the area.³⁷

While it is not possible to know the exact number, Amnesty International has estimated that, throughout the years of the conflict, hundreds –potentially thousands – of ethnic Tigrayans have been detained by the government.³⁸ Most of the people detained were imprisoned without charge or access to a lawyer.³⁹ Thousands of Tigrayans were held inside informal detention centres, including private warehouses and military camps, while some media outlets reported that some detainees passed away while in these detention centres.⁴⁰

A Tigrayan woman living in Bahir Dar, Abrehet Kahsay, told Amnesty International of her experience of detention in December 2020:⁴¹

“I went to a bus station, but then when I went there a group of men, several men, came and asked me for ID. First I asked them why because they were normal people, in normal clothes, so they didn’t look like policemen. They said, ‘some of us are police and we suspected you.’”

Despite living in Bahir Dar for 10 years and having an ID from the city, the group of men identified Abrehet as a member of the Tigrayan community based on her name, and she was detained by the police without any explanation. At the police station, Abrehet was forced to hand over her personal items:

“They checked all my private things. Luckily, they didn’t delete my academic files...I was not even allowed to contact my family or lawyer. Just only to talk to one person. So, when I asked them ‘can I have

³¹ Access Now, “Two years of internet shutdowns: people in Tigray, Ethiopia, deserve better”, 4 November 2022, <https://www.accessnow.org/press-release/two-years-internet-shutdowns-tigray/>

³² OCHA, *Ethiopia: Tigray Region Humanitarian Update*, Situation Report No.4, 20 November 2020, <https://reliefweb.int/report/ethiopia/ethiopia-tigray-region-humanitarian-update-situation-report-no-4-20-november-2020>

³³ UNHCHR, “Internet shutdowns: trends, causes, legal implications and impacts on a range of human rights”, A/HRC/50/55, 13 May 2022,

³⁴ Amnesty International & Human Rights Watch, *We will erase you from this land: Crimes against humanity and ethnic cleansing in Ethiopia’s Western Tigray Zone* (previously cited).

³⁵ Amnesty International & Human Rights Watch, *We will erase you from this land: Crimes against humanity and ethnic cleansing in Ethiopia’s Western Tigray Zone* (previously cited).

³⁶ Amnesty International & Human Rights Watch, *We will erase you from this land: Crimes against humanity and ethnic cleansing in Ethiopia’s Western Tigray Zone* (previously cited).

³⁷ Amnesty International & Human Rights Watch, *We will erase you from this land: Crimes against humanity and ethnic cleansing in Ethiopia’s Western Tigray Zone* (previously cited).

³⁸ Amnesty International, “Ethiopia: Tigrayans targeted in fresh wave of ethnically motivated detentions in Addis Ababa”, 12 November 2021, <https://www.amnesty.org/en/latest/news/2021/11/ethiopia-tigrayans-targeted-in-fresh-wave-of-ethnically-motivated-detentions-in-addis-ababa>

³⁹ Amnesty International, “Ethiopia: Tigrayans targeted in fresh wave of ethnically motivated detentions in Addis Ababa” (previously cited).

⁴⁰ Reuters, “In Ethiopia’s civil war, thousands of jailed Tigrayans endured squalor”, 17 June 2022, <https://www.reuters.com/investigates/special-report/ethiopia-conflict-prisoners>

⁴¹ Amnesty International interview by video call with Abrehet Kahsay, 19 April 2023.

a contact number from my phone?', they said no – I remembered only one number, so I called my friend, and she came and helped me a lot. Actually, I am alive because of her help.”⁴²

Although Abrehet was fortunate that her friend was able to help her out of detention, her experience of being targeted based on her ethnicity was far from uncommon. Amnesty International found that in November 2021, Tigrayans living in Addis Ababa were arrested through house-to-house searches by security forces with the support of vigilante groups conducting ID checks of people on the street.⁴³ People detained by these groups were held in youth recreation centres and other informal detention sites because the police stations in Addis Ababa were at capacity.⁴⁴ Across Ethiopia, Tigrayans lived in fear that they could be rounded up and forcibly disappeared at any time.⁴⁵ International human rights law prohibits arbitrary detention, enforced disappearance and discrimination in all circumstances.

Other human rights organizations, such as Human Rights Watch, also found that Tigrayans arrested during this period appeared to have been targeted on the basis of their ethnicity.⁴⁶ Some people in detention were moved from Addis Ababa to the Afar region, over 200km away, while the whereabouts of others remained unknown after their transfer from the city.⁴⁷ The arrest campaign followed government abuses against Tigrayans based in Addis Ababa, which started after the outbreak of conflict in Tigray in November 2020, after which time authorities in the city arbitrarily arrested and dismissed Tigrayans working in the government and civil service from work, profiled Tigrayans during systematic ID checks, and in some cases repeatedly searched homes without warrants.⁴⁸

In December 2021, the UN Human Rights Council established an International Commission of Human Rights Experts on Ethiopia (hereafter “ICHREE” or “Commission”), tasking it with investigating allegations of violations and abuses of international human rights law, humanitarian law and refugee law in Ethiopia, committed since 3 November 2020 by all parties to the conflict.⁴⁹

In June 2022, the ICHREE delivered a statement to the media, emphasising the importance of continued international attention on Ethiopia, noting that “[t]he ongoing spread of violence, fuelled by hate speech and incitement to ethnic-based and gender-based violence, are early warning indicators of further atrocity crimes against innocent civilians, especially women and children.”⁵⁰

In September 2022, the Commission concluded that it found “reasonable grounds to believe that the parties to the conflict in Ethiopia have committed serious violations and abuses of international human rights law and international humanitarian law.”⁵¹ The report further found that the violations were of such severity that there are reasonable grounds to believe that, in several instances, the violations amount to war crimes and crimes against humanity.⁵² The Commission found that many of the indicators and triggers contained in the United Nations framework of analysis for atrocity crimes were present in Ethiopia, including the dissemination of “hate speech,” which the Commission noted showed “no signs of abating” and lack of independent mechanisms to address it.⁵³ The Commission recommended that all parties to the conflict

⁴² Amnesty International interview by video call with Abrehet Kahsay, 19 April 2023.

⁴³ Amnesty International, “Ethiopia: Tigrayans targeted in fresh wave of ethnically motivated detentions in Addis Ababa” (previously cited).

⁴⁴ Amnesty International, “Ethiopia: Tigrayans targeted in fresh wave of ethnically motivated detentions in Addis Ababa” (previously cited).

⁴⁵ OHCHR, “Press briefing notes on Ethiopia”, 16 November 2021, <https://www.ohchr.org/en/press-briefing-notes/2022/01/press-briefing-notes-ethiopia?LangID=F&NewsID=27816>

⁴⁶ Human Rights Watch, “Ethiopia: Ethnic Tigrayans forcibly disappeared”, 18 August 2021, <https://www.hrw.org/news/2021/08/18/ethiopia-ethnic-tigrayans-forcibly-disappeared>

⁴⁷ Human Rights Watch, “Ethiopia: Ethnic Tigrayans forcibly disappeared” (previously cited).

⁴⁸ Human Rights Watch, “Ethiopia: Ethnic Tigrayans forcibly disappeared” (previously cited).

⁴⁹ UNHRC, International Commission of Human Rights Experts on Ethiopia, <https://www.ohchr.org/en/hr-bodies/hrc/ichre-ethiopia/index#:~:text=In%20its%20resolution%20S-33%2F1%2C%20on%20the%20situation%20of,the%20work%20undertaken%20by%20the%20joint%20investigative%20team>

⁵⁰ OHCHR, “Statement to the media by the International Commission of Human Rights Experts on Ethiopia: delivered by Commission Chair Kaari Betty Murungi and Commission Member Steven Ratner”, Geneva, 30 June 2022, <https://www.ohchr.org/en/statements/2022/06/statement-media-international-commission-human-rights-experts-ethiopia-delivered>

⁵¹ OHCHR, “International Commission of Human Rights Experts on Ethiopia finds reasonable grounds to believe that the federal government has committed crimes against humanity and that Tigrayan forces have committed serious human rights abuses, some amounting to war crimes”, 22 September 2022, <https://www.ohchr.org/en/news/2022/09/international-commission-human-rights-experts-ethiopia-finds-reasonable-grounds>

⁵² OHCHR, “International Commission of Human Rights Experts on Ethiopia finds reasonable grounds to believe that the federal government has committed crimes against humanity and that Tigrayan forces have committed serious human rights abuses, some amounting to war crimes”

⁵³ OHCHR, “International Commission of Human Rights Experts on Ethiopia finds reasonable grounds to believe that the federal government has committed crimes against humanity and that Tigrayan forces have committed serious human rights abuses, some amounting to war crimes” (previously cited).

immediately cease activities that may trigger atrocity crimes, explicitly mentioning the dissemination of “hate speech.”⁵⁴

Warning of the risk of further atrocity crimes, Commission member Radhika Coomaraswamy said:

“Violations and abuses during the conflict show that atrocities do not occur in a vacuum and are directly linked to ethnically motivated hate speech and hostility. Such sentiments have not only been allowed to flourish unchecked, they have been actively promulgated by state officials and the media.”⁵⁵

3.3 THE FACEBOOK PLATFORM’S IMPORTANCE IN ETHIOPIA’S INFORMATION ECOSYSTEM

Ethiopia’s information ecosystem⁵⁶ began to change significantly as internet penetration exploded in the country; the country went from having the lowest proportion of internet users in the world in 2011, with 1.1% internet penetration, to 20.6% internet penetration in 2021.⁵⁷

As internet penetration increased, online news media became a popular source of information. Facebook quickly became the most popular social media platform in Ethiopia, with many internet users using it as a news source, and to share information and opinions.⁵⁸ The limited availability of varied media sources “turned Facebook into an important platform for politicians, activists, journalists and academics to target a still limited, but increasingly broad, audience, and to reach out to younger Ethiopians.”⁵⁹ Activists took up the alternative space created online for activism. The popular group of bloggers known as Zone9 were among those who used Facebook to campaign for democracy and justice before the government arrested most of the members and charged them with terrorism.⁶⁰ As more Ethiopians have come online, Facebook usage has also increased. In 2017, it was estimated that there were 3.3 million Facebook users in the country.⁶¹ In 2021, this estimate had risen to 6.4 million Facebook users, approximately 5.3% of the Ethiopian population.⁶² In April 2020, before the outbreak of armed conflict in northern Ethiopia, the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression noted the importance of the Facebook platform to Ethiopia’s information ecosystem and advised that “[Meta]...deliver localized support to its burgeoning user base to ensure that its platform contributes to people’s expressions, rather than becoming a tool for the spread of hate and disinformation.”⁶³

Despite a relatively low internet penetration rate, the Ethiopian political opposition and diaspora have been able to leverage the internet, and social media specifically, to great advantage. The popular protest that led to the resignation of Prime Minister Hailemariam Desalegn⁶⁴ and the appointment of Abiy Ahmed as a prime minister, widely known as the Oromo Protest⁶⁵, was mainly led by the diaspora using Facebook. The significant influence that social media can have in Ethiopia was further demonstrated in October 2019 by Jawar Mohammed – discussed in more detail in Chapter 5.

⁵⁴ OHCHR, “International Commission of Human Rights Experts on Ethiopia finds reasonable grounds to believe that the federal government has committed crimes against humanity and that Tigrayan forces have committed serious human rights abuses, some amounting to war crimes” (previously cited).

⁵⁵ OHCHR, “UN experts warn of potential for further atrocities amid resumption of conflict in Ethiopia”, 19 September 2022, <https://www.ohchr.org/en/press-releases/2022/09/un-experts-warn-potential-further-atrocities-amid-resumption-conflict>

⁵⁶ The term ‘information ecosystem’ in this context refers to the media and telecommunication landscape of the country, and the sources people use to receive news and information.

⁵⁷ DataReportal, “Digital 2021: Ethiopia”, 11 February 2021, [https://datareportal.com/reports/digital-2021-ethiopia#:~:text=There%20were%2023.96%20million%20internet,%25\)%20between%202020%20and%202021.](https://datareportal.com/reports/digital-2021-ethiopia#:~:text=There%20were%2023.96%20million%20internet,%25)%20between%202020%20and%202021.)

⁵⁸ Iginio Gagliardone, et al., “Mechachal: Online debates and elections in Ethiopia. From hate speech to social engagement on social media”, 2016, <https://ora.ox.ac.uk/objects/uuid:da10c4ee-2726-41cf-ba21-72c9f7cbc440>

⁵⁹ Iginio Gagliardone, et al., “Mechachal: Online debates and elections in Ethiopia. From hate speech to social engagement on social media” (previously cited).

⁶⁰ BBC: “Ethiopia Zone 9 bloggers charged with terrorism”, 18 July 2014, <https://www.bbc.com/news/world-africa-28366841>

⁶¹ DataReportal, “Digital 2017: Ethiopia”, 1 February 2017, <https://datareportal.com/reports/digital-2017-ethiopia?rq=Ethiopia>

⁶² DataReportal, “Digital 2021: Ethiopia”, 11 February 2021, <https://datareportal.com/reports/digital-2021-ethiopia?rq=Ethiopia>

⁶³ UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Report: *Visit to Ethiopia*, 29 April 2020, UN Doc. A/HRC/44/49/Add.1, para. 37.

⁶⁴ BBC: “Ethiopia PM Hailemariam Desalegn in surprise resignation”, 15 February 2018, <https://www.bbc.com/news/world-africa-43073285>

⁶⁵ The Conversation: “Why the Oromo protests mark a change in Ethiopia’s political landscape”, 14 August 2016, <https://theconversation.com/why-the-oromo-protests-mark-a-change-in-ethiopias-political-landscape-63779>

Amnesty International has previously warned of the human rights risks associated with Meta's use of content-shaping algorithms, and the particular risk they pose in conflict-affected settings.⁶⁶

These algorithms are designed to increase an individual's engagement on the Facebook platform, prompting them to stay longer on the platform, increasing their exposure to the tracking-based ads which are Meta's primary source of revenue. Additionally, the longer a person spends on Facebook the more personal data the platform collects about them, in turn better targeting them with ads. This business model has proven to be incredibly profitable for Meta and has led to the company constantly seeking new ways to gather people's intimate personal data. This business model – characterized by ubiquitous surveillance for the mass harvesting of users' data in exchange for the 'free' use of its platforms – is referred to as the 'surveillance-based' business model.⁶⁷

Evelyn Douek, an expert in content moderation, explained to Amnesty International the importance of considering algorithms in understanding how harmful content spreads across the platform:

“If you think about content moderation too narrowly, as only the rules and take downs and leave ups, you are missing the most important decisions a platform makes about how content is presented and circulated.”⁶⁸

Whistle-blower Frances Haugen highlighted the risks associated with the Facebook platform's use of content-shaping algorithms in her 2021 US Senate testimony – particularly the way in which the algorithms boost harmful, inflammatory content in order to keep people engaged.⁶⁹ The Facebook Papers, which formed the basis of Haugen's disclosures, offer a detailed insight into what the company knew about the risks of its business model in conflict-affected settings, and showcases how there was a lack of consideration for these risks when considering changes to the platform – even after the company's contribution to severe human rights harms against the Rohingya in Myanmar in 2017, through its algorithms proactively amplifying and promoting content which incited violence, hatred and discrimination against the community.⁷⁰

⁶⁶ Amnesty International, *Surveillance Giants: How the business model of Google and Facebook threatens human rights* (previously cited); Amnesty International, *The Social Atrocity: Meta and the right to remedy for the Rohingya* (previously cited).

⁶⁷ Amnesty International, *Surveillance Giants: How the business model of Google and Facebook threatens human rights* (previously cited).

⁶⁸ Amnesty International interview by voice call with Evelyn Douek, 12 April 2023.

⁶⁹ MIT Technology Review, “The Facebook whistleblower says its algorithms are dangerous. Here's why”, 5 October 2021, <https://www.technologyreview.com/2021/10/05/1036519/facebook-whistleblower-frances-haugen-algorithms/>

⁷⁰ Amnesty International, *The Social Atrocity: Meta and the right to remedy for the Rohingya* (previously cited).

4. LEGAL FRAMEWORK

4.1 BUSINESS AND HUMAN RIGHTS STANDARDS

Under international law, states have an obligation to respect, protect and fulfil human rights. The UN Guiding Principles on Business and Human Rights (UN Guiding Principles) explicitly outline that such a duty applies in the context of corporate activities.⁷¹

According to the same framework, all companies have a responsibility to “do no harm”. They should respect human rights, regardless of their size, sector, or where they operate.⁷² This 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.⁷³

For business enterprises to respect human rights, “they should avoid infringing on the human rights of others and should address adverse human rights impacts with which they are involved.”⁷⁴

The responsibility to protect human rights requires that business enterprises “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 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.”⁷⁵ According to the OECD Due Diligence Guidelines, an enterprise “contributes” to an impact if its activities, in combination with the activities of other entities cause, facilitate or incentivize another entity to cause an adverse impact.⁷⁶ Contribution must be substantial, meaning it doesn’t include minor or trivial contributions.⁷⁷

Additionally, the UN Guiding Principles stipulate that to meet the 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 this due diligence, a business enterprise might identify that it may contribute – or is already contributing to – human rights abuses. If such a finding occurs, the business enterprise must prevent or cease the negative human rights impacts.⁷⁸

Business enterprises should “treat the risk of causing or contributing to gross human rights abuses as a legal compliance issue wherever they operate.”⁷⁹ Additional guidance from the UN Working Group on Business and Human Rights on operating in conflict-prone regions states that “heightened human rights due

⁷¹ 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, https://www.ohchr.org/sites/default/files/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf, Principle 1.

⁷² 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, <https://www.oecd.org/daf/inv/mne/48004323.pdf>

⁷³ UN Guiding Principles, Principle 11 including Commentary.

⁷⁴ UN Guiding Principle 11.

⁷⁵ UN Guiding Principles, Principle 13 including Commentary.

⁷⁶ OECD Due Diligence Guidance for Responsible Business Conduct, 2018, <https://mneguidelines.oecd.org/OECD-Due-Diligence-Guidance-for-Responsible-Business-Conduct.pdf>, p. 70.

⁷⁷ OECD Due Diligence Guidance for Responsible Business Conduct, 2018, <https://mneguidelines.oecd.org/OECD-Due-Diligence-Guidance-for-Responsible-Business-Conduct.pdf>, p. 70.

⁷⁸ UN Guiding Principles, Commentary to Principle 19.

⁷⁹ UN Guiding Principle 23.

diligence” should incorporate conflict-sensitivity and atrocity-prevention approaches in order to account for the two-way interaction between the business 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.⁸⁰ The Working Group also establishes that when operating in conflict-affected settings the business responsibility to remediate human rights harms should also include engagement with transitional justice processes, for example prosecution initiatives, truth-seeking processes, reparations programmes and institutional reform.⁸¹ The Working Group report makes specific mention of technology companies, explaining that “there is no exceptionalism” and that “the [technology] sector should adopt a genuine human rights approach, in which all rights are recognized as equal, rather than the misguided understanding whereby the right to free speech, or the right to physical security, would be so absolute and unyielding as to trump any other human rights.”⁸²

To verify whether adverse human rights impacts are being addressed, business enterprises should track the effectiveness of their response.⁸³ Tracking should “be based on appropriate qualitative and quantitative indicators” and “draw on feedback from both internal and external sources, including affected stakeholders.”⁸⁴

Transparency is a key component of human rights due diligence. The UN Guiding Principles make clear that companies need to “know and show that they respect human rights” and “showing involves communication, providing a measure of transparency and accountability to individuals or groups who may be impacted and to other relevant stakeholders.”⁸⁵

4.2 HUMAN RIGHTS DUE DILIGENCE AND SOCIAL MEDIA’S ALGORITHMIC TECHNOLOGIES

The UN Guiding Principles provide an important and relevant standard which Big Tech companies should follow, including the need to conduct due diligence on their algorithmic technologies. In 2020, the UN Office of the High Commissioner for Human Rights’ (OHCHR) B-Tech project outlined the relevance of the UN Guiding Principles for technology companies, stating that:

“The UNGPs set out a principled approach for all companies – regardless of industry sector, size, structure or operating context – to identify risks to people and to take action to prevent or mitigate them. This includes the expectation that technology companies make efforts to anticipate and mitigate harms that might occur related to the use of their products or services.”⁸⁶

OHCHR’s B-Tech project specifies that tech companies’ due diligence processes must also include addressing situations in which “business model-driven practices and design decisions create and exacerbate human rights risks,” and an analysis that looks at the unique human rights risks posed by different products and services, end users, and use contexts.⁸⁷ Additionally, “substantive standards for artificial intelligence systems” set out by the Special Rapporteur on the right to freedom of expression specifies that “[c]ompanies should orient their standards, rules and system design around universal human rights principles...”⁸⁸

Furthermore, in 2021 OHCHR set out recommendations for assessing the risks related to artificial intelligence, which included:⁸⁹

⁸⁰ Working Group on the issue of human rights and transnational corporations and enterprises, “Business, human rights and conflict-affected regions: towards heightened action”, 21 July 2020, A/75/212

⁸¹ Working Group on the issue of human rights and transnational corporations and enterprises, “Business, human rights and conflict-affected regions: towards heightened action” (previously cited).

⁸² Working Group on the issue of human rights and transnational corporations and enterprises, “Business, human rights and conflict-affected regions: towards heightened action” (previously cited), para. 99.

⁸³ UN Guiding Principles, Principle 20.

⁸⁴ UN Guiding Principles, Principle 20.

⁸⁵ UN Guiding Principles, Commentary to Principle 21.

⁸⁶ UNOHCHR, “The UN Guiding Principles in the Age of Technology: A B-Tech Foundational Paper”, <https://www.ohchr.org/sites/default/files/Documents/Issues/Business/B-Tech/introduction-ungp-age-technology.pdf>

⁸⁷ UNOHCHR, “Addressing Business Model Related Human Rights Risks: A B-Tech Foundational Paper”, August 2020, https://www.ohchr.org/sites/default/files/Documents/Issues/Business/B-Tech/B_Tech_Foundational_Paper.pdf

⁸⁸ Special Rapporteur on the promotion and protection of the right to freedom of 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

⁸⁹ OHCHR, “The right to privacy in the digital age”, 15 September 2021, A/HRC/48/31.

- a) Systematically conduct human rights due diligence through 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 the AI are, the more transparency is needed;
- c) Ensure participation of all relevant stakeholders 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.

The European Union (EU) has been a leader in terms of attempting to regulate social media companies and algorithmic technologies, passing the Digital Services Act (DSA) in 2022.⁹⁰ The DSA includes an obligation for Very Large Online Platforms (VLOPs) to assess and mitigate systemic risks that arise from the “design, including algorithmic systems, functioning and use made of their services.”⁹¹ Providers of VLOPs will be subject to yearly independent audits to assess compliance with their due diligence obligations.⁹²

In a 2022 report on disinformation and freedom of expression during armed conflicts, the UN Special Rapporteur on the right to freedom of expression specifically mentioned the importance of social media companies conducting heightened due diligence in these settings, stating:

“Companies should carry out heightened human rights due diligence and trigger enhanced risk management strategies in a timely way for pre-, post-, or ongoing conflicts with adequate resources, language and contextual expertise, and engagement of civil society. Due diligence processes should incorporate robust analysis of the impact of companies’ operations, products and services, including the business model itself, on conflict dynamics as well as human rights.”⁹³

4.3 THE CORPORATE RESPONSIBILITY TO PROVIDE REMEDY

Access to remedy is a key pillar of the business and human rights framework. The UN Guiding Principles stipulate 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.”⁹⁴ When a business enterprise finds “that they have caused or contributed to adverse impacts, they should provide for or cooperate in their remediation through legitimate processes.”⁹⁵ Potential impacts – or human rights risks – require action to prevent harm or mitigate the risks as far as possible and to the extent which it may do so. It is therefore impossible for any business enterprise to meet the responsibility to respect human rights if they contribute to human rights abuses and fail to meaningfully remedy the adverse impact.⁹⁶

⁹⁰ Amnesty International, *What the EU's Digital Services Act means for human rights* (Index: POL 30/5830/2022), 7 July 2022, <https://www.amnesty.org/en/documents/pol30/5830/2022/en/>

⁹¹ Amnesty International, *What the EU's Digital Services Act means for human rights* (previously cited).

⁹² Amnesty International, *What the EU's Digital Services Act means for human rights* (previously cited).

⁹³ Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Irene Khan, *Report on Disinformation and freedom of opinion and expression during armed conflicts*, 12 August 2022, A/77/288, para. 124.

⁹⁴ UN Guiding Principles, Principle 22.

⁹⁵ UN Guiding Principles, Principle 22.

⁹⁶ OHCHR, “Frequently Asked Questions about the UN Guiding Principles on Business and Human Rights”, (UN Guiding Principles FAQ), 2014, <https://www.ohchr.org/en/publications/special-issue-publications/frequently-asked-questions-about-guiding-principles>, Question 35, p. 36.

4.4 THE PROHIBITION OF ADVOCACY OF HATRED UNDER INTERNATIONAL HUMAN RIGHTS LAW

Under international human rights law, 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.⁹⁷ However, the right to freedom of expression is not absolute and it can be restricted under certain circumstances, including when it is necessary and proportionate to protect the rights of others.

At the time of writing, Meta's hate speech policy 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 serious disease", adding, "[w]e 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 to offline violence."⁹⁸

While there is no universally agreed definition of "hate speech," international human rights law and standards provide important guidance. Article 20 of the International Covenant on Civil and Political Rights (ICCPR) requires states to prohibit any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence.⁹⁹ Article 4 of the International Convention on the Elimination of All Forms of Racial Discrimination (ICERD) also contains a prohibition of advocacy of hatred.¹⁰⁰ Additionally, under Article 7 of the Universal Declaration of Human Rights (UDHR), everyone has the right to be free from incitement to discrimination.¹⁰¹

The right to equality and non-discrimination, a critical component of international human rights law, constitutes a "basic and general principle relating to the protection of human rights,"¹⁰² and individuals who have their right to non-discrimination violated must have access to effective remedy. This is affirmed by the Toronto Declaration – a civil society-led statement based on international human rights law outlining principles of this fundamental right in the use of machine learning and artificial intelligence.¹⁰³ The declaration states:

"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."¹⁰⁴

As Article 20 of the ICCPR makes clear, 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. Laws prohibiting advocacy of hatred must also comply with the ICCPR's provisions on the right to freedom of expression, and in particular must meet the requirements of legality, legitimate aim, necessity and proportionality.

The Rabat Plan of Action on the prohibition of national, racial or religious hatred constituting incitement to discrimination, hostility or violence suggests a six-part threshold test to guide states' implementation of the prohibition of advocacy of hatred. The six factors that should be considered when determining if an expression constitutes advocacy of hatred are context, the speaker's position or status, intent, content and form, the extent of the speech act and the likelihood of – including imminence – of harm.¹⁰⁵

⁹⁷ UN Human Rights Committee, General Comment 34, CCPR/C/GC/34, para. 11.

⁹⁸ Meta, "Hate Speech Policy", <https://transparency.fb.com/en-gb/policies/community-standards/hate-speech/>

⁹⁹ International Covenant on Civil and Political Rights (ICCPR), Article 20.2.

¹⁰⁰ International Convention on the Elimination of All Forms of Racial Discrimination (ICERD), Article 4.

¹⁰¹ Universal Declaration on Human Rights, <https://www.un.org/en/about-us/universal-declaration-of-human-rights>

¹⁰² United Nations Human Rights Committee, General Comment No.18, UN.Doc. RI/GEN/1/Rev.9 Vol.I (1989), para. 1.

¹⁰³ Amnesty International and Access Now, *The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems* (Index: POL 30/8447/201817), 17 May 2018, <https://www.amnesty.org/es/documents/pol30/8447/2018/en/>

¹⁰⁴ Amnesty International and Access Now, *The Toronto Declaration* (previously cited), para. 53.

¹⁰⁵ United Nations Human Rights Council, Rabat Plan of Action, A/HRC/22/17/Add.4, 11 January 2013, para. 29.

The Rabat Plan of Action emphasizes that tackling the root causes of intolerance requires a much broader set of policy measures, including education, pluralism and diversity, and policies empowering minorities and Indigenous Peoples to exercise their right to freedom of expression.¹⁰⁶ This is due to the central role that nurturing a culture of peace, tolerance and mutual respect in society has in tackling advocacy of hatred. The Rabat Plan of Action notes that “States, media and society have a collective responsibility to ensure that acts of incitement to hatred are spoken out against and acted upon with appropriate measures, in accordance with international human rights law.”¹⁰⁷

The Rabat Plan of Action also 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 which may justify a civil suit; and forms of expression that are not criminal but still raise concerns in terms of tolerance, civility and respect for the convictions of others.¹⁰⁸

For the purposes of this report, the use of the term “advocacy of hatred” refers to expression that constitutes incitement to discrimination, hostility or violence that must be prohibited in law in accordance with ICCPR Article 20. 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 determinations about whether specific pieces of content on the Facebook platform should be considered “advocacy of hatred.” Rather, this report is intended to provide an analysis of Meta’s overall contribution to human rights abuses against the Tigrayan community during the northern Ethiopia conflict, due to its failure to adequately mitigate the risks of the Facebook platform. This includes the algorithmic amplification of content targeting the Tigrayan community, which in some cases may have reached the threshold for advocacy of hatred.

¹⁰⁶ Rabat Plan of Action (previously cited), para. 37.

¹⁰⁷ Rabat Plan of Action (previously cited), para. 35.

¹⁰⁸ Rabat Plan of Action (previously cited), para. 20.

5. FANNING THE FLAMES: META'S ROLE IN STOKING ETHNIC TENSIONS IN ETHIOPIA

“Facebook played a major role in spreading hateful speech, incitement targeting Tigrayans including verbal harassment, videos and graphic photographs... I can say Facebook has fuelled ethnic violence in the Ethiopian civil war.”

Abrham Meareg, Tigrayan community member.¹⁰⁹

This section outlines the role that the Facebook platform played in the spread of hateful content, including advocacy of hatred constituting incitement to violence, hostility and discrimination, against the Tigrayan community and other ethnic groups in Ethiopia both before and during the armed conflict that broke out in Tigray at the end of 2020 and then spread to the Afar and Amhara regions.

5.1 THE ROLE OF FACEBOOK IN OFFLINE VIOLENCE IN ETHIOPIA BEFORE THE ARMED CONFLICT

Despite Ethiopia's relatively low internet penetration rate, Facebook is an important part of the information ecosystem and the content which circulates on the platform can have an impact in the offline world. This became clear in two key incidents which took place one year and five months before the outbreak of the conflict in northern Ethiopia, respectively.¹¹⁰

A clear demonstration of the Facebook platform's significant impact is the response to opposition figure Jawar Mohammed's posts in 2019.¹¹¹ Jawar is an influential Oromo political party leader and activist who was the face of the protests against Ethiopia's government between 2015 and 2018, specifically in Oromia,

¹⁰⁹ Amnesty International interview by video call with Abrham Meareg, 11 April 2023.

¹¹⁰ Special Rapporteur on the promotion and protection of the right of freedom of expression, David Kaye, 'Report on visit to Ethiopia', 23 June 2020, A/HRC/44/49/Add.1, para. 37.

¹¹¹ Human Rights Watch, "Ethiopia: Justice needed for deadly October violence", 1 April 2020, <https://www.hrw.org/news/2020/04/01/ethiopia-justice-needed-deadly-october-violence>

known as the Oromo Protest. He was also one of the founders of Oromia Media Network (OMN). At the time of writing, he had 2.3 million followers on Facebook.¹¹²

In October 2019, Jawar posted on the platform that he believed his life was in danger as the government allegedly wanted to withdraw police protection at his house.¹¹³ The post immediately ignited widespread protests across Oromia regional state, with 1,000 supporters gathering outside of Jawar's residence.¹¹⁴ Posts inciting violence were circulated on Facebook, including a fake video which included a claim that a local official was arming young men.¹¹⁵

After two days of protests during which time 86 people were killed by security forces and in intercommunal violence, Jawar returned to Facebook calling for an end to the violence, posting: "Open the blocked roads, clean the towns of barricades, treat those who have been injured during the protests and reconcile with those you have quarrelled with."¹¹⁶ The violence ceased shortly after.

In June 2020, Oromo singer and activist Hachalu Hundesa was killed by three individuals a group, whose motives for the killing remain unclear.¹¹⁷ In the aftermath, the platform saw a surge in calls for "revenge" particularly aimed at the Amhara ethnic group, leading to a brutal wave of mob violence which, together with a police crackdown, resulted in the killing of at least 177 people.¹¹⁸ On 30 June 2020, in a disproportionate move that would have had adverse impacts on a range of human rights, the Ethiopian government shut down the internet, ostensibly in an attempt to quell the tide of hatred and incitement to violence.¹¹⁹ The shutdown lasted three weeks. Soon after this incident, Meta translated its community standards into Amharic and Oromo for the first time.¹²⁰

The violence that erupted after the killing of Hachalu Hundesa was a moment where many Ethiopian activists and human rights defenders saw the risks that the Facebook platform presented:

"The day that news came out [about the murder of Hachalu Hundesa] Facebook became a platform for violence. There were open calls to commit violence against certain groups, some coded words like 'settlers', coded words for Amhara. Within hours these calls were amplified by [users of] Facebook."

Fisseha Tekle, Amnesty International researcher and human rights defender.¹²¹

Befekadu Hailu, Executive Director of Ethiopian digital rights organization the Centre for the Advancement of Rights and Democracy (CARD), which monitored social media during the conflict in northern Ethiopia, told Amnesty International that "We noticed that when the conflict intensified on the ground, disinformation and hate speech campaigns also intensified on social media."¹²²

Analysis conducted by the Bureau of Investigative Journalism and The Observer found that Facebook was riddled with posts inciting violence against Tigrayans.¹²³ Just one example cited was of a social media influencer calling for people to "cleanse" the Amhara territories of "junta", a term used to refer to Tigrayan forces and later on to target the Tigrayan community.¹²⁴

¹¹² Jawar Mohammed official Facebook page, <https://www.facebook.com/Jawarmd/>

¹¹³ Jawar Mohammed, Facebook post on 23 October 2019, <https://www.facebook.com/Jawarmd/posts/10104932193431623>

¹¹⁴ Reuters, "Protests spread after stand-off at Ethiopian activist's home", 23 October 2019, <https://www.reuters.com/article/us-ethiopia-politics/protests-spread-after-stand-off-at-ethiopian-activists-home-idUSKBN1X211D>

¹¹⁵ BBC News, "Ethiopia violence: Facebook to blame, says runner Gebrselassie", 2 November 2019, <https://www.bbc.co.uk/news/world-africa-50276603>

¹¹⁶ Reuters, "Ethiopia activist calls for calm after 16 killed in clashes", 24 October 2019, <https://www.reuters.com/article/us-ethiopia-politics/hundreds-gathered-outside-house-of-ethiopian-activist-after-deadly-day-of-protests-idUSKBN1X3118>; Yahoo, "Ethiopia activist calls for calm after 16 killed in clashes", 24 October 2019, <https://news.yahoo.com/hundreds-gathered-outside-house-ethiopian-093356262.html>; Human Rights Watch, "Ethiopia: Justice needed for deadly October violence", 1 April 2020, <https://www.hrw.org/news/2020/04/01/ethiopia-justice-needed-deadly-october-violence>

¹¹⁷ BBC, "Hachalu Hundessa: Ethiopia's murdered musician who sang for freedom", 2 July 2020, <https://www.bbc.com/news/world-africa-53238206>

¹¹⁸ Amnesty International, 'Ethiopia: Account for all people arrested after Hachalu Hundesa's killing', 18 July 2020.

¹¹⁹ Nation, "Ethiopia restores full internet access after 3 weeks of shutdown", 23 July 2020, <https://nation.africa/kenya/news/africa/ethiopia-restores-full-internet-access-after-3-weeks-of-shutdown-1905118>; UNHRC, "Internet shutdowns: trends, causes, legal implications and impacts on a range of human rights", A/HRC/50/55, 13 May 2022.

¹²⁰ Council on Foreign Relations, "Facebook's Content Moderation Failures in Ethiopia", 19 April 2022, <https://www.cfr.org/blog/facebook-content-moderation-failures-ethiopia>

¹²¹ Amnesty International interview by video call with Fisseha Tekle, 19 April 2023.

¹²² Amnesty International interview by video call with Befekadu Hailu, 24 July 2023.

¹²³ The Bureau for Investigative Journalism, "Facebook accused by survivors of letting activists incite ethnic massacres with hate and misinformation in Ethiopia", 20 February 2022, <https://www.thebureauinvestigates.com/stories/2022-02-20/facebook-accused-of-letting-activists-incite-ethnic-massacres-with-hate-and-misinformation-by-survivors-in-ethiopia>

¹²⁴ The Bureau for Investigative Journalism, "Facebook accused by survivors of letting activists incite ethnic massacres with hate and misinformation in Ethiopia" (previously cited).

Freweyni Itsay, a member of the Tigrayan community, shared her view that calls to violence against her community were widely shared on the Facebook platform after the conflict broke out in northern Ethiopia:

“Calling for the arrest of people, calling for the killing of people, claiming that a Tigrayan father killed his own son because his mother was from another ethnic group and such things were being promoted on Facebook.”¹²⁵

A number of media outlets and organizations have documented some of the vast quantity of harmful content including content targeting the Tigrayan community and inciting violence, hostility and discrimination, which circulated on Facebook after the outbreak of the conflict in northern Ethiopia in 2020.¹²⁶ The International Commission of Human Rights Experts on Ethiopia (ICHREE) found that narratives insinuating that all Tigrayans were supporting the “enemy” were prevalent on social media.¹²⁷ This content was posted by a variety of actors, including the Prime Minister of Ethiopia, prominent government-aligned activists, government-aligned media outlets, groups and other accounts with large followings.

Lucy Kassa, a journalist who was based in Ethiopia throughout the beginning of the conflict, explained the impact of these posts on the Tigrayan community to Amnesty International:

“There are always hate comments related to ethnicity, it was always like that. But what made it grave for the Tigrayans was that the Ethiopian government was systematically targeting them. They were being rounded up in Addis and in the rest of the country and sent to detention camps because of their ethnicity, and they were being called ‘junta’.”¹²⁸

Lucy Kassa, journalist

The term “junta” was flagged by many of the Tigrayan community members interviewed by Amnesty International as a coded word, which was used to imply that all Tigrayan people either supported the TPLF or were TPLF members themselves. Fisseha Tekle also explained his personal view on the use of the term to Amnesty International:

“[The word] ‘junta’ was first used by the Prime Minister, and it was used to mean any Tigrayan... it was used against everyone. Everyone believed to be Tigrayan was labelled as a ‘junta’, even in ordinary conversations, ordinary language, so it was on social media too.”¹²⁹

Amnesty International has also previously recorded the word “junta” being used against Tigrayans in the Western Tigray Zone by Amhara Special Forces (ASF), Fano militia, and people belonging to other ethnic groups.¹³⁰

Timnit Gebru, an AI ethicist and founder and executive director of the Distributed AI Research Institute (DAIR), who also follows social media trends in Ethiopia, described how content targeting the Tigrayan community on the Facebook platform intensified as the conflict went on:

“The scariest time was around October 2021 – there was like, basically people saying with urgency that ‘you should kill them all now, even if you feel bad about it you should do it now... Other stuff was very widespread... talking about ‘cleaning up’ and ‘weeding out the pests’ and ‘cancer’ and ‘daytime hyena.’”¹³¹

¹²⁵ Amnesty International interview by video call with Freweyni Itsay, 12 May 2023.

¹²⁶ Vice, “How Facebook is stoking a civil war in Ethiopia”, 8 November 2021, <https://www.vice.com/en/article/qjbpd7/how-facebook-is-stoking-a-civil-war-in-ethiopia#:~:text=On%20Oct,accounts%20before%20Facebook%20removed%20it.>; Council on Foreign Relations, “Facebook’s content moderation failures in Ethiopia” (previously cited); The Guardian, “Facebook fails test to detect violent hate speech in global ads – again”, 9 June 2022, <https://www.theguardian.com/us-news/2022/jun/09/facebook-hate-speech-test-fail-meta>

¹²⁷ UN Human Rights Council, *Report of the International Commission of Human Rights Experts on Ethiopia*, 14 September 2023, A/HRC/54/55, para. 33.

¹²⁸ Amnesty International interview by video call with Lucy Kassa, 7 July 2023.

¹²⁹ Amnesty International interview by video call with Fisseha Tekle, 19 April 2023.

¹³⁰ Amnesty International & Human Rights Watch, *We will erase you from this land: Crimes against humanity and ethnic cleansing in Ethiopia’s Western Tigray Zone* (previously cited).

¹³¹ Amnesty International interview with Timnit Gebru, 7 July 2023.

5.2 THE ROLE OF PUBLIC FIGURES AND GOVERNMENT ALLIES IN SPREADING CONTENT TARGETING THE TIGRAYAN COMMUNITY

Abiy Ahmed is the first Ethiopian Prime Minister who is active on social media and at the time of writing, has 4.5 million followers on the Facebook platform.¹³² This is perhaps indicative of the increasing importance of social media in Ethiopia's information ecosystem, and of an understanding among political actors that platforms such as Facebook are key battlegrounds for shaping the narrative in times of crisis.

In 2021, Amnesty International noted an alarming rise in social media posts advocating ethnic violence – including posts inciting violence and using ethnic slurs against Tigrayans – and raised concerns about government officials imploring citizens to take up arms against both the TPLF and the Oromo Liberation Army (OLA).¹³³ Additionally, the ICHREE reported that many instances of incitement to discrimination, hostility and violence were found on social media throughout the conflict, including on the Facebook platform.¹³⁴

At times, Abiy Ahmed himself posted content targeting the Tigrayan community. For example, in July 2021, he wrote a Facebook post vowing to crush the “cancerous” rebels, whom he described as “weeds” – a term frequently used by government-backed accounts to describe Tigrayans.¹³⁵ As of October 2023, the post remains available on the Facebook platform.

This Facebook post was written in a context where the term “cancer” and other insulting or derogatory language, such as “*junta*”, “*woyane*” and “*agame*” were increasingly used to incite discrimination, hostility and violence against the Tigrayan community.¹³⁶ Anti-TPLF narratives evolved over time to include Tigrayan people in general. The insinuation that all Tigrayan people supported the TPLF was made by Abiy Ahmed himself, in a 30 June 2021 statement, in which he implied that all Tigrayans were supporting the “enemy”.¹³⁷ Even before the outbreak of the armed conflict in northern Ethiopia, Abiy Ahmed had referred to TPLF leaders as “daytime hyenas”, a term which was used to describe the Tigrayan community.¹³⁸ In the same month, Meta removed a post by Abiy Ahmed which urged Ethiopian citizens to rise up and “bury” Tigray forces who were moving towards the country's capital. In the deleted post, Abiy wrote that “the obligation to die for Ethiopia belongs to all of us” and he called on citizens to mobilize by “holding any weapon or capacity”.¹³⁹

A content moderator who worked on the Facebook platform's Ethiopia content during the armed conflict explained to Amnesty International the influence that Abiy Ahmed's words had on the platform:

“There was a time when the Prime Minister used the term ‘*ye Ken jib*’ which is ‘daytime hyenas’ and after that reference, [other] Ethiopians started using that word for Tigrayans. He kept on using the word ‘*junta*’ and afterwards other Ethiopians... would use the word to refer to Tigrayans.”¹⁴⁰

Taye, former Meta content moderator

Pro-government activists also used the Facebook platform to incite violence and discrimination against Tigrayans. In October 2021, Dejene Assefa, an activist with over 120,000 followers on Facebook, called for Ethiopians to take up arms against the Tigrayan community, writing: “The war is with those you grew up

¹³² Abiy Ahmed Facebook page, https://www.facebook.com/PMAbiyAhmedAli/?locale=en_GB

¹³³ Amnesty International, “Ethiopia: Sweeping emergency powers and alarming rise in online hate speech as Tigray conflict escalates”, 5 November 2021, <https://www.amnesty.org/en/latest/news/2021/11/ethiopia-sweeping-emergency-powers-and-alarming-rise-in-online-hate-speech-as-tigray-conflict-escalates/>

¹³⁴ International Commission of Human Rights Experts on Ethiopia, “Comprehensive findings and legal determinations”, 13 October 2023, <https://www.ohchr.org/sites/default/files/documents/hrbodies/hrcouncil/chreethiopia/a-hrc-54-crp-3.pdf>, para. 451.

¹³⁵ Abiy Ahmed, Facebook post on 18 July 2021, <https://www.facebook.com/112704996810839/posts/573861097361891/>

¹³⁶ UN Human Rights Council, *Report of the International Commission of Human Rights Experts on Ethiopia*, 14 September 2023, A/HRC/54/55, para. 33.

¹³⁷ The Times, “Abiy's pledge to crush weeds of Tigray raises fears of genocide in Ethiopia”, 19 July 2021, <https://www.thetimes.co.uk/article/abiys-pledge-to-crush-weeds-of-tigray-raises-fears-of-genocide-in-ethiopia-sx0xkb2v>; UN Human Rights Council, *Report of the International Commission of Human Rights Experts on Ethiopia*, 14 September 2023, A/HRC/54/55, para. 3.

¹³⁸ Financial Times, “Ethiopian ethnic rivalries threaten Abiy Ahmed's reform agenda”, 27 March 2019, <https://www.ft.com/content/1c8aac04-457f-11e9-a965-23d669740bfb>

¹³⁹ The Guardian, “Facebook fails test to detect violent hate speech in global ads – again” (previously cited).

¹⁴⁰ Amnesty International interview by video call with Taye (pseudonym, real name withheld for security reasons), 22 May 2023.

with, with your neighbour. If you can rid your forest of these thorns...victory will be yours.”¹⁴¹ The post was shared nearly 900 times and attracted 2,000 reactions before it was removed.

All of the Tigrayan community members Amnesty International spoke to referred to the impact that posts of this nature had offline. Abrehet Kahsay recalled seeing a change in people’s behaviour towards the Tigrayan community living outside of the Tigray region:

“People from... the federal government, they were spreading false narratives. There was all this hate speech on social media, and they were aggravating it. Regular citizens were acting weird. They assumed every Tigrayan was a criminal, so it was very bad. It was very bad especially for people living in the Amhara region because the tensions were between the federal government and Tigray and Tigrayans and Amharas.”¹⁴²

5.3 ANALYSIS: HOW ADVOCACY OF HATRED TARGETING THE TIGRAYAN COMMUNITY LINKED TO OFFLINE VIOLENCE

As detailed in the preceding section, messages advocating hatred against Tigrayans, as well as other dehumanizing and discriminatory content targeting the Tigrayan community, became prevalent on the Facebook platform during the conflict in northern Ethiopia. This section provides an analysis of the connection between this online content and offline acts of violence.

The UN’s 2014 Atrocity Prevention Framework lists 14 risk factors and indicators which allow the manifestation of atrocity crimes, such as genocide, war crimes, crimes against humanity and ethnic cleansing.¹⁴³ Enabling circumstances and preparatory action are listed as a common risk factor for atrocity crimes within the framework. These are defined as “[e]vents or measures, whether gradual or sudden, which provide an environment conducive to the commission of atrocity crimes, or which suggest a trajectory towards their preparation.”¹⁴⁴ “Increased inflammatory rhetoric, propaganda campaigns or messages advocating hatred against protected groups, populations or individuals” is considered an indicator that this risk is present.¹⁴⁵ The comment to this risk factor notes that such actions can “create an environment that favours or even encourages the commission of [atrocity] crimes.”¹⁴⁶

On 20 July 2021, Alice Wairimu Nderitu, the UN Special Adviser on the Prevention of Genocide released a statement condemning inflammatory statements made by political leaders and armed groups in Ethiopia, saying:

“The use of pejorative and dehumanizing language like “cancer”, “devil”, “weed” and “bud” to refer to the Tigray conflict is of utmost concern. Hate speech, together with its propagation through social media is part of a worrisome trend that contributes to further fuel ethnic tensions in the country.”¹⁴⁷

Academic literature on genocide and other forms of mass violence draws a connection between advocacy of hatred and dehumanizing language and acts of mass violence. With regard to dehumanization as a justification for violence, Adam Jones, author of *Genocide: A Comprehensive Introduction*, notes that “before 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.”¹⁴⁸ In the Ethiopian context, dehumanizing terms such as “weeds” and “daytime hyena” were frequently used to refer to Tigrayans on social media. Jones highlights the centrality of the “less dramatic, permitted, everyday acts of

¹⁴¹ Rest of World, “Why Facebook keeps failing in Ethiopia”, 13 November 2021, <https://restofworld.org/2021/why-facebook-keeps-failing-in-ethiopia/>

¹⁴² Amnesty International interview by video call with Abrehet Kahsay, 19 April 2023.

¹⁴³ OHCHR, *Framework for analysis for Atrocity Crimes: a tool for prevention*, 30 September 2014, <https://www.ohchr.org/en/documents/tools-and-resources/framework-analysis-atrocity-crimes>

¹⁴⁴ OHCHR, *Framework for analysis for atrocity crimes: A tool for prevention* (previously cited), p. 16.

¹⁴⁵ OHCHR, *Framework for analysis for atrocity crimes: A tool for prevention* (previously cited).

¹⁴⁶ OHCHR, *Framework for analysis for atrocity crimes: A tool for prevention* (previously cited), p. 16.

¹⁴⁷ Statement by Alice Wairimu Nderitu, United Nations Special Adviser on the Prevention of Genocide, on the continued deterioration of the situation in Ethiopia, 20 July 2021, <https://reliefweb.int/report/ethiopia/statement-alice-wairimu-nderitu-united-nations-special-adviser-prevention-genocide>

¹⁴⁸ Adam Jones, *Genocide: A Comprehensive Introduction*, 2nd edition, 2010, p. 393.

violence that make participation (under other conditions) in genocidal acts possible”, which include, “all expressions of social exclusion, dehumanization, depersonalization, pseudo-specification, and reification that normalize... violence toward others.”¹⁴⁹ Jones’ emphasis on the importance of what becomes “permitted” in a society in the lead up to atrocities is supported by Mahmood Mamdani, who advises that in order to understand the dynamics which allow for mass killings to occur, “we need to move away from an assumption of genocide as a conspiracy from above and to examine how both “perceptions” and “circumstances” alter the intent and motivations of perpetrators.”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 posits 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 Social media platforms enable the dissemination of ideological narratives on an unprecedented scale, and this can present a particular risk in fragile and conflict-affected settings. Scholar Edward Kissi, writing in a comparative study of mass violence in Ethiopia, Rwanda and Cambodia during the 1970s, noted that ideology was a common thread in all contexts “...ideology and revolution created the conditions for the destructions of groups in Rwanda, Ethiopia and Cambodia, [while] the scope, pace and success of murder depended on the degree of control and influence that the perpetrators exercised in their respective society.”¹⁵²

The phenomenon of “epistemic dependence” is another way of understanding how individuals can be influenced to engage in acts of violence against a targeted group. It highlights the importance of the presence of trusted sources of information, including media sources and civic leaders, which legitimise and promote violence and dehumanisation. Maynard writes that “atrocities-justifying ideologies are most influential when they operate through such epistemic dependence.”¹⁵³ Befekadu Hailu, the Executive Director of the Centre for the Advancement of Rights and Democracy (CARD), an Ethiopian civil society organization which focuses on digital rights, explained to Amnesty International the perception of the Facebook platform as an important source of information in Ethiopia:

“It’s very important because we don’t have strong media institutions or civil society organisations. Therefore, it has filled a gap – it is the place where journalists get information, it is a place where activists and human rights defenders network, it is a place where advocacy activities take place... All of these things were not possible without Facebook in particular, because it is the most widely used social media.”¹⁵⁴

Freweyni Itsay, a Tigrayan community member living in Addis Ababa, explained how she felt that people considered information seen on the Facebook platform to be legitimate:

“Some of my colleagues thought that whatever was on Facebook was right, whatever was being said about Tigrayans was correct. There were a lot of people taking in the hate, people thought whatever was written was right.”¹⁵⁵

Lucy Kassa, an Ethiopian journalist, shared her view that content on Facebook impacted the way the Tigrayan community was viewed during the northern Ethiopia conflict:

“Okay, the government is targeting Tigrayans, but why would you target your neighbour? It’s because on Facebook there were activists dehumanising the Tigrayans and saying you shouldn’t trust them. That’s the reason neighbours were reporting [their Tigrayan neighbours] to the police.”¹⁵⁶

¹⁴⁹ Adam Jones, *Genocide: A Comprehensive Introduction* (previously cited), p. 437.

¹⁵⁰ Mahmood Mamdani, *When Victims Become Killers: Colonialism, Nativism and the Genocide in Rwanda*, 2001.

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 Edward Kissi, “Rwanda, Ethiopia and Cambodia: links, faultlines and complexities in a comparative study of genocide”, *Journal of Genocide Research*, 2004, 6:1, 115 – 133, p. 116.

¹⁵³ 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. 827.

¹⁵⁴ Amnesty International interview by video call with Befekadu Hailu, 24 July 2023.

¹⁵⁵ Amnesty International interview by video call with Freweyni Itsay, 12 May 2023.

¹⁵⁶ Amnesty International interview by video call with Lucy Kassa, 7 July 2023.

Discussing the contribution of Meta's algorithms to the spread of content targeting Tigrayans and inciting violence and discrimination on the Facebook platform, Timnit Gebru, an AI ethicist and the executive director of DAIR told Amnesty International:

“There were a lot of explicit calls for killing and exterminating people. Even the most explicit ones – even after they were reported they would stay up for days and that combined with the algorithm maximizing engagement, that means they would be shared a lot more widely.”¹⁵⁷

The spread of ideology is central to the creation of an environment that allows for the outbreak of mass violence. Ideology can be spread through narratives of dehumanization, impending threat from the “other”, or false information about wrongs supposedly committed by the group. As this ideology becomes normalized, it may influence the decision-making of policy makers, provide direction to the perpetrators of violent acts, and create support for violent actions against the targeted group. In the context of the violence perpetrated against the Tigrayan community, the narratives equating Tigrayan civilians with TPLF forces played a central role in enabling and justifying acts of violence perpetrated by soldiers, militia groups and civilians.

This connection between content targeting the Tigrayan community 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 Tigrayan community. As has been demonstrated, the mass dissemination of messages that advocated hatred inciting violence and discrimination against the Tigrayans, as well as other dehumanizing and discriminatory content targeting the community, was auxiliary in creating an enabling environment for mass violence against the target group, with severe consequences for a range of human rights.

5.4 CONTENT MODERATION AND SPEECH TARGETING TIGRAYANS ON THE FACEBOOK PLATFORM

“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 Ethiopian context amid rising levels of inter-ethnic tensions before 2020, and after the beginning of the northern Ethiopia conflict.

Content moderation involves the detection of harmful content such as “hate speech” (messages advocating hatred) and then a range of varied actions may be taken in response, such as the removal of content, making the content less visible (known as “demoting” or “downranking”), or penalties against individuals or groups posting such content.¹⁵⁸ Content moderation is conducted with a combination of human content moderators and content moderation automated AI systems.¹⁵⁹

Meta has for many years had internal policies against the use of “hate speech” on its platform. The company's community standards recognize that “hate speech” on the platform can lead to offline violence and other harms, stating: “We don't allow hate speech on Facebook. It creates an environment of intimidation and exclusion, and in some cases may promote offline violence.”¹⁶⁰

In 2021, former Meta employee and whistle-blower Frances Haugen leaked a large number of internal Meta documents – known as the ‘Facebook Papers’ –, which contained revelations about the company's approach to protecting platform users and told the US Senate that the company prioritized “profits over safety”.¹⁶¹ The Facebook Papers showed that Meta's content moderation systems failed to adequately and effectively detect divisive content in non-English speaking countries such as Ethiopia, including in those that Meta had deemed to be at “high risk” of violence.¹⁶² There has been extensive media coverage of allegations that

¹⁵⁷ Amnesty International interview by video call with Timnit Gebru, 7 July 2023.

¹⁵⁸ Amnesty International, *The Social Atrocity: Meta and the right to remedy for the Rohingya* (previously cited).

¹⁵⁹ Amnesty International, *The Social Atrocity: Meta and the right to remedy for the Rohingya* (previously cited).

¹⁶⁰ Meta, ‘Hate Speech’, <https://transparency.fb.com/en-gb/policies/community-standards/hate-speech/>; Note: Meta's definition of “hate speech” differs from the definition of “advocacy of hatred” under Article 20 of the ICCPR, with Meta's definition including a broader range of expression which would not always reach the threshold of Article 20.

¹⁶¹ United States Senate Committee on Commerce, Science and Transportation, ‘Statement of Frances Haugen’, 4 October 2021.

¹⁶² Financial Times, “You can't unsee it: the content moderators taking on Facebook”, 18 March 2023, <https://www.ft.com/content/afeb56f2-9ba5-4103-890d-91291aea4caa>

Meta's efforts to curb harmful content in African countries such as Ethiopia were woefully inadequate, particularly when compared to efforts to monitor content in countries considered more politically relevant.¹⁶³

The importance of content moderation practices reflecting social and linguistic diversity was highlighted to Amnesty International by Zecharias Zelalem, an Ethiopian journalist who covered the conflict in northern Ethiopia and the Facebook platform's role in disseminating messages that advocated hatred and incited violence and discrimination. He told Amnesty International that he believes the way language was used changed during the northern Ethiopia conflict:

“There are certain terms in the Amharic language that are not a normal part of the vocabulary, but thanks to government activists, terms that were used to demonise [certain] populations were normalised through Facebook.”¹⁶⁴

Many of the affected individuals Amnesty International spoke to highlighted the frequency with which they saw the terms “junta” and “daytime hyena” being used to describe Tigrayans on the Facebook platform.¹⁶⁵

One of the key revelations of the Facebook Papers was the existence of the “tier system” used by Meta to rank the world's countries for the purposes of investment in content moderation.¹⁶⁶ Countries within different “tiers” are given vastly different content moderation resources – and they are organised into tiers based on information which is not public or subject to external review.¹⁶⁷

The Facebook Papers also show that Meta has long known that its content moderation systems in at-risk countries have been inadequate. In June 2020 – four months before the northern Ethiopia conflict began – employees reviewing the platform's “signals” – the data collected from platform users and partners which are used to understand problematic content – said that they found “significant gaps” in the most at-risk countries, including Myanmar and Ethiopia, “showcasing that our current signals may be inadequate.”¹⁶⁸ Like in Myanmar, the Facebook platform's automated content removal system could not read the local typeface in Ethiopia.¹⁶⁹ Additionally, further internal documents from 2020 reveal that Meta did not have a hate speech classifier – an algorithm trained to automatically detect key terms which indicate a post may be advocating hatred – in Ethiopia, despite its status as an “at-risk country”.¹⁷⁰ Digital rights expert and prominent human rights defender Befekadu Hailu discussed his view that Meta's moderation of Ethiopian content was inadequate during the conflict, saying:

“some content that should have been removed; it was not removed. In general, the moderation of content related to the conflict was done with poor human resource and analysis, and delays [in content removal]”¹⁷¹

This was echoed in the final report of the ICHREE in October 2023, which noted that “credible sources indicated to the Commission that they found multiple other instances where harmful content was flagged multiple times was not removed.”¹⁷²

¹⁶³ Rest of World, “AI content moderation is no match for hate speech in Ethiopian languages”, 27 June 2023, <https://restofworld.org/2023/ai-content-moderation-hate-speech/>; Cybernews, “Pressure increases for Meta to overhaul African content moderation”, 7 February 2023, <https://cybernews.com/editorial/pressure-increases-for-meta-to-overhaul-african-content-moderation/>; VOA Africa, “Meta called to boost Africa content moderation”, 24 January 2023, <https://www.voaafrica.com/a/meta-called-to-boost-africa-content-moderation/6931579.html>; Reuters <https://cybernews.com/editorial/pressure-increases-for-meta-to-overhaul-african-content-moderation/>; “Meta urged to boost African content moderation as contractor quits”, 24 January 2023, <https://www.reuters.com/world/africa/meta-urged-boost-africa-content-moderation-contractor-quits-2023-01-24/>; Addis Insight, “Facebook's content moderation in Ethiopia is under-funded and under-resourced”, 25 June 2022, <https://addisinsight.net/facebook-content-moderation-in-ethiopia-underfunded-and-under-resourced/>; The Guardian, “Facebook ‘lets vigilantes in Ethiopia incite ethnic killing’”, 20 February 2022, <https://www.theguardian.com/technology/2022/feb/20/facebook-lets-vigilantes-in-ethiopia-incite-ethnic-killing>; CNN Business, “Facebook has language blind spots around the world which allow hate speech to flourish”, 26 October 2021, <https://edition.cnn.com/2021/10/26/tech/facebook-papers-language-hate-speech-international/index.html>

¹⁶⁴ Amnesty International interview by video call with Zecharias Zelalem, 20 April 2023.

¹⁶⁵ Amnesty International interview by video call with Abrehet Kahsay, 19 April 2023; Amnesty International interview by video call with Fisseha Tekle, 19 April 2023; Amnesty International interview by video call with Lucy Kassa, 7 July 2023.

¹⁶⁶ The Verge, “The tier list: how Facebook decides which countries need protection”, 25 October 2021, <https://www.theverge.com/22743753/facebook-tier-list-countries-leaked-documents-content-moderation>

¹⁶⁷ The Verge, “The tier list: how Facebook decides which countries need protection” (previously cited).

¹⁶⁸ Rest of World, “Why Facebook keeps failing in Ethiopia” (previously cited).

¹⁶⁹ The Verge, “The tier list: how Facebook decides which countries need protection” (previously cited).

¹⁷⁰ The Verge, “The tier list: how Facebook decides which countries need protection” (previously cited).

¹⁷¹ Amnesty International interview by video call with Befekadu Hailu, 24 July 2023.

¹⁷² International Commission of Human Rights Experts on Ethiopia, “Comprehensive findings and legal determinations”, 13 October 2023, <https://www.ohchr.org/sites/default/files/documents/hrbodies/hrcouncil/chreethiopia/a-hrc-54-crp-3.pdf>, para. 453.

Documents in the Facebook Papers written as late as 2021 acknowledge that, though many of the countries which Meta refers to as “At Risk” speak multiple language and dialects – including Ethiopia –, moderation teams are equipped to handle only some of these languages and a large amount of “hate speech” slips through the net to remain on the platform.¹⁷³ Frances Haugen detailed how this meant that Meta’s attempts at content moderation were often too slow and reactive in crisis situations, with the company sometimes training a new language model after a crisis has already occurred:

“The most fragile places in the world are linguistically diverse places, and they speak languages that are not spoken by tons of people. They add a new language usually under crisis conditions.”¹⁷⁴

This approach is demonstrated in the Facebook Papers, which show that the company knew it didn’t have sufficient coverage in local languages to actively identify messages that advocated hatred in Ethiopia.¹⁷⁵ In 2020, internal company documents acknowledged that Meta had insufficient moderation capabilities in Amharic, a language that is widely spoken across the country and is Ethiopia’s federal working language, and that language barriers also stopped platform users in Ethiopia from reporting problematic content.¹⁷⁶ An internal document from 2021 listed more than a dozen languages across the Facebook and Instagram platforms that were “prioritized” for expanding its automated systems during the first half of that year – the languages had been partly prioritized based on the risk of “offline violence” – Amharic and Afan Oromo were included in the list.¹⁷⁷ However, Ethiopia’s armed conflict had been underway for several months at that point. Furthermore, in a separate internal document from 2021, the company found Ethiopia to be at “dire” risk of violence and in its assessment of the company’s response to violent and inciting content, it ranked its own capacity in the country as a zero out of three.¹⁷⁸ In another document, a Meta staff member acknowledged that the company lacked “human review capacity” for Ethiopia in the run up to the country’s elections.¹⁷⁹

Taye, a Facebook content moderator working for SamaSource, Inc. (Sama), a company contracted by Meta to conduct content moderation for Africa, who moderated Ethiopian content during the conflict, explained how the lack of human review capacity meant that content moderation processes were sluggish, particularly when moderators flagged new terms being used to incite violence against Tigrayans on the platform:

“We would get very slow responses, from Facebook themselves. Because how the line of communication worked was [that] we reported to people who were quality analysts...and then they report to people in Ireland, the market specialists. But after that is reported, the feedback we would get – it could take months...”¹⁸⁰

In a letter to Amnesty International dated 18 September 2023, Meta stated that they can now review content in the top four languages spoken in Ethiopia (Amharic, Oromo, Somali and Tigrinya).¹⁸¹ Meta also highlighted that in 2021 they increased investment in proactive detection technology in Ethiopia and that they hired more Ethiopian nationals to work on content moderation.¹⁸² While these improvements are welcome, it remains to be seen if Meta has increased its content moderation capacity enough to adequately mitigate harms in Ethiopia. Moreover, content-based solutions cannot alone be considered an adequate mitigation for the algorithmic amplification of harmful content.

¹⁷³ CNN Business, “Facebook has language blind spots around the world which allow hate speech to flourish” (previously cited).

¹⁷⁴ CNN Business, “Facebook has language blind spots around the world which allow hate speech to flourish” (previously cited).

¹⁷⁵ Rest of World, “Why Facebook keeps failing in Ethiopia” (previously cited).

¹⁷⁶ Rest of World, “Why Facebook keeps failing in Ethiopia” (previously cited); Reuters, “Meta urged to boost African content moderation as contractor quits” (previously cited); CNN Business, “Facebook has language blind spots around the world which allow hate speech to flourish” (previously cited).

¹⁷⁷ CNN Business, “Facebook has language blind spots around the world which allow hate speech to flourish” (previously cited).

¹⁷⁸ Wired, “A New Lawsuit Accuses Meta of Inflaming Civil War in Ethiopia”, 13 December 2022, <https://www.wired.com/story/meta-hate-speech-lawsuit-ethiopia/>

¹⁷⁹ Wired, “A New Lawsuit Accuses Meta of Inflaming Civil War in Ethiopia” (previously cited).

¹⁸⁰ Amnesty International interview by video call with Taye, 22 May 2023.

¹⁸¹ Meta letter to Amnesty International, 18 September 2023.

¹⁸² Meta letter to Amnesty International, 18 September 2023

5.5 CONTENT MODERATION: THE HUMAN COST

Social media platforms are increasingly utilizing specially trained algorithms to conduct automated content moderation. Content moderation algorithms can both flag content for human review and take automated decisions on pieces of content, such as by “demoting” the content in the News Feed or removing content altogether.¹⁸³ Many critics of Meta’s approach to content moderation cite a need for more human oversight – due to issues discussed above such as the need to understand local contexts and cultures.¹⁸⁴ However, it is important to note that there is a high human cost to content moderation, particularly for moderators based in Global Majority countries such as Kenya, where Meta’s content moderation for Ethiopia takes place, due to the poor working conditions many moderators endure.

In 2019, Meta contracted with Sama, a company operating in Nairobi labelling data to train AI software for clients such as Meta and Tesla, to take on the work of content moderation.¹⁸⁵ Sama had not performed content moderation work before but accepted the contract and hired people from across the African continent to work on content moderation. Sama employed around 25 content moderators to moderate content from Ethiopia – a country with a population of 117 million people – in three of the 85 languages spoken in the country – Amharic, Tigrinya and Oromo.¹⁸⁶

In March 2023, 184 former content moderators brought a case against both Sama and Meta for alleged human rights violations and wrongful termination of contracts. In August 2023, all parties to the case agreed to mediation, with a view to settling out of court.¹⁸⁷ All the content moderators involved in the litigation complained of constant pressure from managers to work at speed, with a requirement to deal with each “ticket”, or item, in 50 to 55 seconds.¹⁸⁸ Many of the content moderators have reported suffering from mental health issues such as depression, anxiety, post-traumatic stress disorder and suicidal ideation after spending hours a day reviewing content which contained graphic violence.¹⁸⁹

Taye, a former content moderator at Sama, explained the toll that the work has taken on him and his colleagues:

“You just question all the things you knew or thought you knew. And [this] kind of makes you lose your trust and hope in humanity as well...I know a lot of colleagues who don’t trust people anymore because the job makes you see the worst part of human beings.”¹⁹⁰

Many of the content moderators claim that they did not fully understand what they were signing up for when they took the job, with some believing that they would be doing the equivalent of a customer service role.¹⁹¹

A former social media platform moderator, Richard Mathenge, who was also employed by Sama, told Amnesty International that he did not really understand the nature of the role he took up:

“When we got this opportunity, at first it looked very enticing. When time actually defined what we were getting into, it became very frustrating, it became devastating as well...Because of the traumatic conditions that we were subjected to... it became worse and worse.”¹⁹²

Regardless of the platform they are contracted to work for, the nature of the job means moderators will regularly see graphic and harmful content. James Oyange Odhiambo, a former social media content

¹⁸³ Amnesty International, *The Social Atrocity: Meta and the right to remedy for the Rohingya* (previously cited).

¹⁸⁴ Reuters, “Meta urged to boost Africa content moderation as contractor quits” (previously cited); MIT Technology Review, “Facebook needs 30,000 of its own content moderators, says report”, 8 June 2020, <https://www.technologyreview.com/2020/06/08/1002894/facebook-needs-30000-of-its-own-content-moderators-says-a-new-report/>; The Verge, “The tier list: how Facebook decides which countries need protection” (previously cited); CNN Business, “Facebook has language blind spots around the world that allow hate speech to flourish” (previously cited).

¹⁸⁵ The Guardian, “‘A watershed’: Meta ordered to offer mental health care to moderators in Kenya”, 7 June 2023, <https://www.theguardian.com/global-development/2023/jun/07/a-watershed-meta-ordered-to-offer-mental-health-care-to-moderators-in-kenya>

¹⁸⁶ Rebecca Hamilton and Rosa Curling, “Facebook Beware: The ‘Rest of World’ is Hitting Back”, Just Security, 6 February 2023, <https://www.justsecurity.org/84982/facebook-beware-the-rest-of-world-is-hitting-back/>

¹⁸⁷ TechCrunch, “Meta and moderators agree to mediation on labor dispute”, 23 August 2023, <https://techcrunch.com/2023/08/23/meta-and-moderators-agree-to-mediation/>; Reuters, “Content moderators sue Meta over alleged ‘union-busting’ in Kenya”, 20 March 2023, <https://www.reuters.com/technology/content-moderators-sue-meta-over-alleged-union-busting-kenya-2023-03-20/>

¹⁸⁸ Financial Times, “You can’t unsee it: the content moderators taking on Facebook” (previously cited).

¹⁸⁹ The Guardian, “‘A watershed’: Meta ordered to offer mental health care to moderators in Kenya” (previously cited).

¹⁹⁰ Amnesty International interview by video call with Taye, 22 May 2023.

¹⁹¹ The Guardian, “‘A watershed’: Meta ordered to offer mental health care to moderators in Kenya” (previously cited).

¹⁹² Amnesty International interview with Richard Mathenge, former content moderator, 7 June 2023.

moderator who worked in Kenya, spoke to Amnesty International about the pivotal role moderators play in attempting to keep the darkest aspects of the digital world away from platform users:

“Content moderators are gatekeepers of morality... Content moderators are like soldiers who go to war and come back with PTSD...”¹⁹³

Investment in automated content moderation may provide benefits in more efficiently identifying instances of misinformation, “hate speech” and other harmful content from the Facebook platform, and better enable Meta to prevent and mitigate adverse human rights impacts. However, content moderation is a complex challenge, and as this chapter demonstrates, automated content moderation has often allowed harmful content to remain online in Ethiopia. It remains the case that automated content moderation cannot adequately serve the vast linguistic and cultural diversity in all the countries in which Meta operates. As such, human content moderators remain crucial to Meta’s business operations. It is therefore imperative that Meta employs a sufficient number of content moderators with the required language skills and contextual understanding, that these individuals are adequately remunerated for their work and have their labour rights respected, and that they are provided with adequate psychosocial support to mitigate the personal impacts of their work.

In a letter to Amnesty International, Sama has said that it disputes the claims made against the company in court and that it is committed to finding a settlement which is agreeable to all parties, and provided details of well being initiatives for content moderators – see the letter in the Annex.¹⁹⁴

¹⁹³ Amnesty International interview with James Oyange Odhiambo, former TikTok content moderator, 7 June 2023.

¹⁹⁴ Sama letter to Amnesty International, 25 October 2023.

6. ONLINE RHETORIC, OFFLINE HARM

“[Meta] don’t accept our thoughts, our concerns to protect our loved ones. As if we don’t have any dreams, as if we don’t have a family to keep alive. I firmly believe if they just took the content down... if they took the reports seriously, my father would be alive.”¹⁹⁵

Abrham Meareg

As outlined above, Meta played a key role in normalizing dehumanizing rhetoric against the Tigrayan community in Ethiopia during the armed conflict. This chapter outlines four cases in which content on the Facebook platform contributed to violence and other harms in the offline world. These cases are not thought to be exhaustive but provide illustrative examples of the nature of the Facebook platform’s role in relation to violence during the conflict in northern Ethiopia.



6.1 THE CASE OF MEAREG AMARE

Abrham Meareg is a petitioner in civil litigation brought against Meta in Kenya, which alleges that the company allowed posts inciting violence to flourish on Facebook – amplified by Meta’s algorithmic recommender systems – and inflaming the conflict in Ethiopia.¹⁹⁶ The petitioners in the case – Abrham Meareg, Fisseha Tekle (an Amnesty International staff member engaging in the litigation in a personal capacity) and the Katiba Institute, a Kenyan-based civil society organisation –, are urging the court to order that Meta takes a number of emergency steps to the spread of harmful violent content on the Facebook platform. These measures include that Meta increases resourcing – including staffing – of its content moderation in Nairobi and creates a restitution fund of USD 2 billion for victims of violence around the world incited by posts on Facebook.¹⁹⁷

Abrham’s father, Professor Meareg Amare, was a well-known and widely respected Tigrayan member of staff at Bahir Dar University and had lived in the city of Bahir Dar for several years. On 9 October 2021, an anonymously-run Facebook page called “BDU Staff”, with over 50,000 followers, posted his picture, announcing he was “hiding” at Bahir Dar University where he was working as a chemistry professor and had

¹⁹⁵ Amnesty International interview by video call with Abrham Meareg, 11 April 2023.

¹⁹⁶ Amnesty International, “Kenya: Meta sued for 1.6 billion USD for fueling Ethiopia ethnic violence”, 14 December 2022, <https://www.amnesty.org/en/latest/news/2022/12/kenya-meta-sued-for-1-6-billion-usd-for-fueling-ethiopia-ethnic-violence/>

¹⁹⁷ Foxglove, “Death by design: a major new case against Facebook”, 14 December 2022, <https://www.foxglove.org.uk/2022/12/14/death-by-design-major-new-case-facebook/>

carried out “abuses”.¹⁹⁸ In the comments, people called for violence against the professor, calling him a “snake” and suggesting that he posed a risk to people from the Amhara ethnic group.¹⁹⁹

The next day, on 10 October 2021, another post was made to the same group. This one contained Amare’s photo as well as the name of the neighbourhood he lived in in Bahir Dar. The post alleged that he had helped to massacre people, that he was a corrupt property owner, that he had helped with TPLF military incursions into nearby areas and that he had stolen huge sums of money.²⁰⁰ Meareg Amare’s widow, Nigist Hailu, told Amnesty International that both she and her husband noticed a change in the way he was treated after the posts were circulated on Facebook:

“As soon as we arrived back home [from where they had been temporarily staying in Addis Ababa], our neighbours who we previously had good relations with, they actually avoided us...”²⁰¹

His son, Abrham, told Amnesty International that he was shocked by the comments on the post about his father:

“There were a lot of comments saying ‘why are you wasting your time writing on Facebook? Go and get him in his house’. They said, ‘why are you warning him by posting on Facebook?’. A lot of the commenters were... encouraging action against him without wasting time.”²⁰²

After being alerted to the posts by a friend, Abrham repeatedly reported them via the Facebook platform from 14 October onwards but was unable to get a response from Meta until 11 November, eight days after his father was killed.

Due to the popularity of the Facebook platform in Ethiopia, Abrham was concerned that if the posts stayed up and continued to be boosted by the platform’s algorithms, there would be a high likelihood of violence against his father:

“I knew Facebook had played a major role in mobilizing people, especially back in 2018, which led to reforms within the EPRDF.²⁰³ So I know how Facebook played a major role, and I knew [the posts] would bring a major consequence on the family. I knew it was a death sentence for my father.”²⁰⁴

On 3 November 2021 – three weeks after the posts appeared on the BDU staff page – a group of men followed Meareg home from the university where he taught and shot him in the legs and the chest outside of his home.²⁰⁵ He lay dying in the street for seven hours, with the men warning onlookers that they would be shot if they tried to give him medical assistance.²⁰⁶

Nigist Hailu, Meareg’s wife, was called by a friend while at church to tell her that her husband had been attacked. She recounted what she saw once she arrived at the scene:

“When I found him I kissed him and called his name but he didn’t respond... The friend who called me said that people had tried to help the professor but they had been told to go away by the attackers... The police told me that the attackers had called the Professor “junta”. ”²⁰⁷

As described in Chapter 5, the word “junta” was frequently used as a term to dehumanize Tigrayan people and had been used by people commenting on the BDU Staff Facebook post about the professor.

¹⁹⁸ NBC News, “Facebook hit with \$2 billion lawsuit connected to political violence in Africa”, 14 December 2022, <https://www.nbcnews.com/tech/misinformation/facebook-lawsuit-africa-content-moderation-violence-rcna61530>

¹⁹⁹ Time, “New lawsuit accuses Facebook of contributing to deaths from ethnic violence in Ethiopia”, 14 December 2022, <https://time.com/6240993/facebook-meta-ethiopia-lawsuit/>; NBC News, “Facebook hit with \$2 billion lawsuit connected to political violence in Africa”, 14 December 2022, <https://www.nbcnews.com/tech/misinformation/facebook-lawsuit-africa-content-moderation-violence-rcna61530>

²⁰⁰ NBC News, “Facebook hit with \$2 billion lawsuit connected to political violence in Africa” (previously cited).

²⁰¹ Amnesty International interview by video call with Nigist Hailu, 23 August 2023.

²⁰² Amnesty International interview by video call with Abrham Meareg, 11 April 2023.

²⁰³ The Facebook platform was used to mobilize anti-government protests in Ethiopia in 2018. The protests led to the EPRDF leader Hailemariam Desalegn to step down and the EPRDF announcing that it would take comprehensive reform measures, including revising and repealing laws that narrowed political expression.

²⁰⁴ Amnesty International interview by video call with Abrham Meareg, 11 April 2023.

²⁰⁵ Foxglove, “Death by design: a major new case against Facebook” (previously cited).

²⁰⁶ Time, “New lawsuit accuses Facebook of contributing to deaths from ethnic violence in Ethiopia” (previously cited).

²⁰⁷ Amnesty International interview by video call with Nigist Hailu, 23 August 2023.

Although Abrham was not present when his father was killed, he believes his father's killers were at least partially motivated by the content of the BDU staff page Facebook post, recounting what he was told by neighbours who witnessed the attack:

“The [killers] were referring to the Facebook post again and again to the people who were trying to administer first aid.”²⁰⁸

Abrham told Amnesty International that following the killing, Amhara militants took possession of the Meareg family home and Abrham's mother fled to Addis Ababa.

Abrham explained to Amnesty International that he was motivated to pursue justice from Meta because he feels that the company has been dismissive towards the concerns of Ethiopians about the spread of hate and incitement to violence on Facebook:

“There are people who lost their livelihood, their dream, due to Facebook's deliberate negligence, and ignorance. And arrogance about accepting our concerns.”²⁰⁹

On 9 November 2021 – 6 days after Meareg Amare was killed, Meta published a blog outlining their efforts to mitigate harm in Ethiopia.²¹⁰

This included measures to reduce possible violating content, removing content calling for violence, and improvements to hate speech enforcement. However, in the case of Professor Meareg Amare, these measures – announced one year after the outbreak of the armed conflict – came far too late. Additionally, according to Abrham Meareg, one of the posts targeting the Professor continued to circulate on the Facebook platform until December 2022.²¹¹

In a September 2023 reply to a letter from Amnesty International asking about the content moderation for their operations in Ethiopia, Meta stated that following a due diligence process conducted in 2021, they have improved and simplified their reporting tools to report possible violating content.²¹² However, it is clear that, even with these improvements, Meta remains unable to deal with harmful content in an adequate and timely manner.



A screenshot of the first post by the BDU staff page targeting Professor Meareg Amare.

²⁰⁸ Amnesty International interview by video call with Abrham Meareg, 11 April 2023.

²⁰⁹ Amnesty International interview by video call with Abrham Meareg, 11 April 2023.

²¹⁰ Meta, 'An update on our our longstanding work to protect people in Ethiopia', 9 November 2021, <https://about.fb.com/news/2021/11/update-on-ethiopia/>

²¹¹ Amnesty International interview by video call with Abrham Meareg, 11 April 2023

²¹² Meta letter to Amnesty International, 18 September 2023.



6.2 THE CASE OF FISSEHA TEKLE

Fisseha Tekle is the second petitioner in the civil litigation against Meta filed in Kenya. Fisseha is an Amnesty International staff member who began to be targeted by both pro-government and anti-government accounts on social media in May 2020 for his research on human rights violations in Ethiopia. His work included documenting Facebook posts, the content of which the litigation alleges led to real-world killings.²¹³ He has been interviewed for this research in his personal capacity.

Fisseha described the way he began to be targeted on Facebook before the outbreak of the armed conflict in 2020, when he reported on human rights violations in Ethiopia:

“People called me a TPLF agent and a TPLF member. They started to associate my name with people from the Tigray region.”²¹⁴

Fisseha noted that he would be targeted by different actors depending on the kind of research he had undertaken:

“When I talked about violations against the Amhara [people], it was Tigrayan [people] coming after me. When I talked about violations against the Tigrayans, it was pro-government Amhara [people].”²¹⁵

Fisseha noted that the peak of the online vitriol against him came after Amnesty International reported on the Axum massacre in February 2021²¹⁶, after the report was leaked before its official publication date:

“Some of the media houses started to talk about me, claiming that I am Tigrayan, that my wife’s family is linked to one of the founders of the TPLF. That content was reproduced on Facebook, and people were sharing it. There were people claiming that I should be hunted.”²¹⁷

When he began to be targeted on Facebook, Fisseha reported the posts and didn’t feel that the platform’s response was adequate:

“There was no response, or after a long time there was a response that it didn’t violate community standards.”²¹⁸

As a result of being constantly targeted on Facebook, Fisseha fears for his safety and has had to change the way he lives his life, avoiding the Ethiopian community in Kenya, where he lives. He believes that it would not be safe for him to return to Ethiopia, mainly due to the Facebook posts inciting violence against him:

“I’m socially disconnected from my family and from my culture and community in general.”²¹⁹



A screenshot of a post targeting Fisseha. It reads: “The renegade’s, the traitor’s and the Junta’s defender Fisseha Tekle.” The post is accompanied by a picture of Fisseha and some text that reads “Amnesty’s fake drama report writer and director, the Junta Fisseha Tekle.”



²¹³ Time, “New lawsuit accuses Facebook of contributing to deaths from ethnic violence in Ethiopia” (previously cited).

²¹⁴ Amnesty International interview by video call with Fisseha Tekle, 19 April 2023.

²¹⁵ Amnesty International interview by video call with Fisseha Tekle, 19 April 2023.

²¹⁶ Amnesty International, *The Massacre in Axum* (Index: AFR 25/3730/2021), 26 February 2021, <https://www.amnesty.org/en/documents/afr25/3730/2021/en/>

²¹⁷ Amnesty International interview by video call with Fisseha Tekle, 19 April 2023.

²¹⁸ Amnesty International interview by video call with Fisseha Tekle, 19 April 2023.

²¹⁹ Amnesty International interview by video call with Fisseha Tekle, 19 April 2023.

The following two case studies, Itsay Assefa and Lucy Kassa, are not connected to the litigation filed in Kenya.



6.3 THE CASE OF ITSAY ASSEFA

Freweyni Itsay is a 25-year-old Tigrayan woman living in Addis Ababa, studying accountancy. Freweyni's family were living in the Oromia region of Ethiopia during the Tigray conflict. In November 2022, her father, Itsay Assefa, was killed by members of the Fano militia after being targeted in Facebook posts, in circumstances similar to those suffered by Meareg Amare. In December 2020, her father was targeted on Facebook by several pages, including news pages such as 'Negarit' and 'Amhara Times', which published posts accusing him of being an Oromo Liberation Army (OLA, sometimes also referred to as OLF-Shene) supporter. In August 2021, OLA announced an alliance with TPLF rebels in the north of Ethiopia.²²⁰

Freweyni explained to Amnesty International that the family were only made aware of the Facebook posts after her father was killed two years later in December 2022, although a family friend later informed the family that Itsay had been aware that the posts were circulating:

"He knew about the posts. He even received a text message with threats on his mobile phone, but he underestimated the posts on Facebook, he never thought it would lead to such events."²²¹

Freweyni was not present during the attack, which took place in the family home, but was told by her family members that the killers referenced the content of the Facebook posts:

"They accused him of supporting the OLA and they killed him. They were mentioning everything that was on Facebook as they were accusing him of these things."²²²

The attackers – who the family believe were associated with the Fano militia in part due to the clothes they wore – also killed Freweyni's younger brother, Dawit.

Freweyni believes that her father was targeted in the posts due to his ethnicity, and she told Amnesty International that her father was not involved in political activities. She described the content of the post and comments inciting violence against her father to Amnesty International:

"It was saying 'people should be aware of these people [Tigrayans] and that the information should be shared with the intelligence services'. Most of the comments were negative, they were saying 'these people will get us killed' and 'we should be worried'. Many others were sharing the post.'" ²²³

Amnesty International reviewed a screenshot of the post (see page 39) from the 'Negarit' news page on Facebook which mentions Itsay Assefa. It reads:

"By sharing this information, let us all make sure it reaches the Ethiopian National Intelligence and Security Service and the Oromia Police Commission.

We have a list of people working to get the Amhara attacked by the OLF-Shene fanning conflict between the Amhara and Oromo people.

We have confirmed that "a feast" of devastation/massacre is in the making in the Anger Gute – a town located in close proximity to Nekemte town in East Wolega Zone, Gida Ayana Woreda; Anger Gute is home to people from various nations and nationalities, particularly to several Amhara and Oromo people.

²²⁰ Al Jazeera, "Ethiopia armed group says it has alliance with Tigray forces", 11 August 2021, <https://www.aljazeera.com/news/2021/8/11/ethiopia-armed-group-says-it-has-alliance-with-tigray-forces>

²²¹ Amnesty International interview by video call with Freweyni Itsay, 12 May 2023.

²²² Amnesty International interview by video call with Freweyni Itsay, 12 May 2023.

²²³ Amnesty International interview by video call with Freweyni Itsay, 12 May 2023.

And yet we have come to learn that there are individuals who are working hard to divide/separate these ‘innocent’ people by taking a mission from Shene. And the individuals with the mission are:

Ato Itsay: Occupation, a businessman who used to own a large farmland during the TPLF regime”

The post then lists the names and occupations of eight other people, before ending with:

“The aforementioned people mainly support Shene and give them tips to escape every time the National Defence Force enters the town”

Freweyni described the impact that losing her father and brother has had on her family:

“It’s very difficult. That they were killed in such a barbaric way. My father was a very good person for his community... My brother was very young. He had a lot of life ahead of him.”²²⁴



Despite Meta’s mitigation measures, the posts targeting Itsay Assefa and others – containing veiled threats of violence – were able to circulate on the Facebook platform for over two years. At the time of Freweyni’s interview with Amnesty International, in May 2023, she reported that some of the posts targeting her father remained on the platform.²²⁵

²²⁴ Amnesty International interview by video call with Freweyni Itsay, 12 May 2023.

²²⁵ Amnesty International interview by video call with Freweyni Itsay, 12 May 2023



6.4 THE CASE OF LUCY KASSA

Lucy Kassa is an Ethiopian freelance journalist who extensively covered the conflict in northern Ethiopia. After each report she released, she was targeted by pro-government and anti-government accounts with posts inciting violence against her on Facebook.²²⁶

Lucy described the way she was targeted both online and offline for her journalism, particularly after reporting on sexual violence in Tigray:

“There were physical attacks and online propaganda, an online campaign by the Ethiopian government – they released a statement [on Facebook].²²⁷ There was a huge campaign, an attempt to discredit my work.”²²⁸

Fearing for her safety, Lucy fled to Nairobi in February 2021, continuing to report on the armed conflict whilst abroad. After reporting on a possible incendiary weapons attack in Tigray in May 2021, Lucy was targeted by a pro-government account with over 200,000 followers, which posted her photo and openly called for her arrest.²²⁹ The post received over 6,000 likes and 1,000 comments.

Many of the comments on the post incited violence against Lucy:

“They were very nasty, misogynistic and calls for attacks. That I should be stoned, and I should be raped...That post was the huge one because it had a lot of likes and shares.”²³⁰

The intensity of the threats online caused Lucy to flee for a second time:

“My photo was on Facebook, the government and influential activists were calling for violence against me, saying that I should be stoned, and the government should “do something” about me. So, it wasn’t safe for me in Nairobi.”²³¹



A screenshot of a Facebook post targeting Lucy Kassa and calling for her to be ‘brought to justice’.



²²⁶ Vice, “How Facebook is stoking a civil war in Ethiopia” (previously cited).

²²⁷ Ethiopia Current Issues Fact Check, Facebook post, 10 February 2021, <https://www.facebook.com/100312305234664/posts/claimed-journalist-is-not-legally-registered-with-the-ethiopian-broadcast-author/153113326621228/>; CFWIJ, “Ethiopia: CFWIJ Condemns The Statement Of The Ethiopian Broadcasting Authority In Response To The Threats Faced By Lucy Kassa”, 12 February 2021, <https://www.womeninjournalism.org/threats-all/ethiopia-cfwij-condemns-the-statement-of-the-ethiopian-broadcasting-authority-in-response-to-the-threats-faced-by-lucy-kassa>

²²⁸ Amnesty International interview by video call with Lucy Kassa, 7 July 2023.

²²⁹ Vice, “How Facebook is stoking a civil war in Ethiopia” (previously cited).

²³⁰ Amnesty International interview by video call with Lucy Kassa, 7 July 2023.

²³¹ Amnesty International interview by video call with Lucy Kassa, 7 July 2023.

Lucy explained to Amnesty International that she was targeted by both pro-government and pro-TPLF accounts online, depending on which human rights violations she was reporting on:

“It would switch based on the story I told. If the story supports their narrative they will support me, but I was basically attacked by all sides, because I was reporting on violations from all sides.” ²³²

Now based in Europe, while she feels physically safe, Lucy still avoids socialising with the Ethiopian community because of the online targeting against her and is coming to terms with the psychological impact of the online hate she experienced:

“At some point I kind of hated my job. I hated doing stories...That was the purpose [of the attacks], so it had an effect. It has affected my mental health.”²³³

These case studies demonstrate the prominent role that the Facebook platform played in relation to human rights harms and offline violence – including violations of the right to life – amid the conflict in Northern Ethiopia. The following chapter explores how Meta’s algorithmic systems and business model facilitated the spread of content which advocated hatred and incited violence in Northern Ethiopia, including in respect of these cases, and analyses Meta’s contribution to human rights harms.

²³² Amnesty International interview by video call with Lucy Kassa, 7 July 2023.

²³³ Amnesty International interview by video call with Lucy Kassa, 7 July 2023.

7. MORE THAN AN ABERRATION: META'S SYSTEMIC FAILURES IN ETHIOPIA

“While in some sense a core component of Facebook is the possibility of going viral, we’re beginning to realize that violating content is more likely to go viral compared to benign.”²³⁴

Unnamed Facebook employee, 2019

7.1 META'S POST-2017 BUSINESS MODEL: ENGAGEMENT BY ANOTHER NAME

Amnesty International has previously found that Meta's surveillance-based business model, based on invasive profiling and targeted advertising, fuels the spread of advocacy of hatred and incitement to violence, hostility and discrimination. In its 2022 report on Meta's contribution to atrocities committed in Myanmar in 2017, the organisation warned that there was a significant risk that the company would again contribute to human rights abuses in conflict-affected settings without wide-ranging reforms to Meta's business practices.²³⁵

In December 2017, just months after the company played a high-profile role in crimes against humanity and the ethnic cleansing of the Rohingya people of Myanmar, Meta pivoted its News Feed algorithm around a new metric, dubbed “meaningful social interactions” (MSI) to rank people's interactions on Facebook. This was enacted as a way to counter declining engagement on the platform (namely, to address the fact that users were commenting, liking, resharing less on the platform).²³⁶ Ranking, in this context, refers to how Meta prioritizes what content users see in their Facebook News Feed, which is decided by content-shaping algorithms.

²³⁴ The Facebook Papers, “Integrity choices and LIKE data”, 13 November 2019, p.4, https://www.documentcloud.org/documents/23605720-tier2_civ_pr_1119

²³⁵ Amnesty International, *The Social Atrocity: Meta and the right to remedy for the Rohingya* (previously cited).

²³⁶ CNN Business, “Likes, anger emojis and RSVPs: the math behind Facebook's News Feed – and how it backfired”, 27 October 2021, <https://edition.cnn.com/2021/10/27/tech/facebook-papers-meaningful-social-interaction-news-feed-math/index.html>

MSI officially launched in 2018, with Mark Zuckerberg branding it part of his efforts to “fix Facebook” and prioritize user well-being.²³⁷ Meta emphasized that posts promoting “meaningful interactions” would be prioritized over content users may see passively. Although the term has not been clearly defined, “meaningful interaction” refers to users’ “closeness” as a proxy for meaningfulness, namely, assuming that a post by a friend would be more “meaningful” to a user than a post by a news source. The intended result was that platform users would see more posts from friends and family than from brands or publishers.²³⁸ However, content which garnered comments and shares would still be prioritized for algorithmic boosting, and virality remained a core feature of the platform.²³⁹

Despite Zuckerberg’s claim that MSI was intended to boost user well-being, it appears that it was intended to further Meta’s business model of maximum engagement for the maximum number of platform users. This is reflected in the Facebook Papers, where the metrics used to consider whether MSI was beneficial include “ecosystem metrics” such as daily active users, the length of time users spent on the platform, revenue, broadcast sharing (broadcasting live video or audio to a large number of people) and reshare broadcast sharing (the sharing of a live broadcast video).²⁴⁰

Although it was hailed as a major change, “meaningful social interactions” appear to have been only very loosely defined, if at all. Speaking at the 2018 Recode Media Conference, then Head of News Feed Adam Mosseri said: “We’re trying to figure out how best to measure and understand that”, suggesting that Meta did not have a clear definition even after MSI’s launch.²⁴¹

An internal Meta memo from 2019 defines MSI as “a reciprocal interaction between at least two users which is characterized by an action from an originator followed by a response from a responder.”²⁴² It is unclear how this definition differs in any significant way from Meta’s previous engagement-based approach, beyond being a tweak to the system. Another issue raised by the lack of definition is that MSI “has a lot of discontinuities and arbitrary-feeling-choices of weights and thresholds,” as noted by a Facebook employee in the Facebook Papers.²⁴³

It also appears from the Facebook Papers that the MSI metric was developed with Global North platform users in mind. An internal note on MSI, written in 2019, outlines that MSI ranking has “been more beneficial to US-CA compared to Global for most [engagement] metrics.”²⁴⁴

Additionally, according to an internal memo, the metric improved user *perception* over the quality of the content they were viewing, suggesting that “[t]his has occurred due to increase of higher rated friend reshare links at the expense of Page links.”²⁴⁵

The amplification of reshared content from friends is a development which could create a perfect storm in polarized settings such as Ethiopia, where Facebook is considered a trustworthy news source. This issue was raised by Meta employees, as demonstrated in the Facebook Papers, with one employee commenting that they were unsure whether the company was valuing MSI generated on reshared content appropriately because “we don’t differentiate between whether the MSI happened on an original post created by a user or whether it happened on a post created by a page that was re-shared by a user.”²⁴⁶ The impact of the lack of differentiation between where an interaction happened means that content posted by pages aiming for virality – often by posting content which is inflammatory, divisive and harmful – would still be amplified by the platform’s algorithms.

The same employee goes on to mention the impact that this lack of differentiation has, outlining that reshares and comments now dictate how much algorithmic amplification a page will receive: “This has had

²³⁷ CNN Business, “Mark Zuckerberg’s goal for 2018: ‘Fixing Facebook’”, 4 January 2018, <https://money.cnn.com/2018/01/04/technology/mark-zuckerberg-2018-goal/index.html?iid=EL>; CNN Business, “Facebook to show more content from friends, less from publishers and brands”, 11 January 2018, <https://money.cnn.com/2018/01/11/technology/facebook-news-feed-change/index.html>

²³⁸ CNN Business, “Facebook to show more content from friends, less from publishers and brands” (previously cited).

²³⁹ CNN Business, “Facebook to show more content from friends, less from publishers and brands” (previously cited).

²⁴⁰ The Facebook Papers, “The Meaningful Social Interactions Metric Revisited: Part 4”, 6 November 2019, https://www.documentcloud.org/documents/21602160-tier0_rank_exp_1119-1, p. 2.

²⁴¹ Recode, “Full interview: Facebook’s head of news partnerships & head of News Feed live from Code Media”, 13 February 2018, <https://www.youtube.com/watch?v=dU-Mlj5vEjU>

²⁴² The Facebook Papers, “The Meaningful Social Interactions Metric Revisited: Part 4” (previously cited), p. 2.

²⁴³ The Facebook Papers, “The Meaningful Social Interactions Metric Revisited: Part 4” (previously cited), p. 10.

²⁴⁴ The Facebook Papers, “The Meaningful Social Interactions Metric Revisited: Part 4” (previously cited).

²⁴⁵ The Facebook Papers, “The Meaningful Social Interactions Metric Revisited: Part 4” (previously cited), p. 18.

²⁴⁶ The Facebook Papers, “The Meaningful Social Interactions Metric Revisited: Part 5”, 6 November 2019, p.4, https://www.documentcloud.org/documents/21602208-tier0_rank_exp_1119-2

some unintended consequences: for example, the capacity for a post to be reshared and garner MSI dominates the ranking of page content.”²⁴⁷ In the same document another employee muses that:

“my intuition tells me that this might cause really significant distortions that hurt users.”²⁴⁸

In a separate internal memo from 2020, a Meta employee voices similar concerns about the type of content users may see and the role of Facebook’s systems: “There are multiple reasons why people have different experiences: for example, our systems are making it worse.”²⁴⁹

Since the launch of MSI in 2018, Meta has adjusted the way interactions were prioritized several times. An internal document – which is undated but likely from 2020²⁵⁰ – shows updated MSI weightings: the lowest weightings are given for an ‘anger’ react, (0), reshare (1.5) and love or care react (2), while the highest weights are given to comments, story replies, story attributed messages and avatar stickers (all weighted at 15).²⁵¹ While the shift away from “reacts” may appear positive, a focus on comments remains potentially problematic as it may incentivize posts which will elicit strong reactions from platform users.

The risk of weighting comments was acknowledged by Meta through the application of guardrails with certain types of comments not receiving weight (and therefore not counting towards the MSI metric), such as: single character comments, comments deleted the same day, comments on engagement bait posts, and comments where the author has already made 250 comments that day.²⁵² However, the document does not refer to any guardrails intended to address comments which are more likely to contain content advocating hatred or inciting violence and discrimination.

7.2 RESHARES AND VIRALITY

Previous analysis by Amnesty International has found that Meta’s use of algorithmic virality – where certain content is amplified to allow it to reach a broader audience than it would have done otherwise – carries significant risks, which Meta has long been aware of, particularly in conflict-affected settings.²⁵³

While the introduction of MSI was mooted to privilege well-being over engagement for engagement’s sake, virality remained a core feature of the Facebook platform, with a particular emphasis on reshares and comments. The Facebook Papers demonstrate that Meta employees were aware that reshares could cause harm in crisis situations, with a document from 2019 outlining an experiment in limiting the number of reshares.²⁵⁴ The employee who ran the experiment reported that it had been successful in reducing a range of harms, including stopping messages that contained calls to violence, but noted that it negatively impacted the MSI metric, “achieving such a win in practice might require a change in the way we formulate and goal on MSI, as the current formulation of MSI is explicitly reduced by a reduction in sharing behaviours despite other core engagement measures being unmoved.”²⁵⁵

The employee recommended that limiting reshares be considered a “break the glass” measure and that:

“a weak form could be considered a pre-crisis or even a steady state change to virality.”²⁵⁶

Documents from the Facebook Papers show that in January 2020, Meta made changes to the MSI metric in an acknowledgment that the virality created by MSI remained a problem on the platform. An internal document states that the goals of the metric are being changed to “create a better friend ecosystem while

²⁴⁷ The Facebook Papers, “The Meaningful Social Interactions Metric Revisited: Part 5” (previously cited), p. 4.

²⁴⁸ The Facebook Papers, “The Meaningful Social Interactions Metric Revisited: Part 5” (previously cited), p. 10.

²⁴⁹ The Facebook Papers, “Auditing Integrity Systems to make sure we don’t hurt racial justice”, 31 August 2020, https://www.documentcloud.org/documents/23605718-tier2_civ_ro_0820, p. 5.

²⁵⁰ The assumption that the document is from 2020 is based on the fact that it contains the ‘weighting’ for different forms of engagement (namely, likes, emojis, comments) for ‘2020H2’.

²⁵¹ The Facebook Papers, “Meaningful Social Interactions Useful Links”, undated, https://www.documentcloud.org/documents/21601160-tier2_rank_other_undated

²⁵² The Facebook Papers, “Meaningful Social Interactions Useful Links” (previously cited), p. 1.

²⁵³ Amnesty International, *The Social Atrocity: Meta and the right to remedy for the Rohingya* (previously cited).

²⁵⁴ The Facebook Papers, “Max Reshare Depth Experiment”, 6 November 2019, https://www.documentcloud.org/documents/21602015-tier1_rank_pr_1119

²⁵⁵ The Facebook Papers, “Max Reshare Depth Experiment” (previously cited), p. 1.

²⁵⁶ The Facebook Papers, “Max Reshare Depth Experiment” (previously cited), p. 5.

reducing virality”,²⁵⁷ remained the most heavily weighted metric, incentivizing shares and replies to comments regardless of the risk that content inciting violence and discrimination may be amplified.

The document from January 2020 shows Meta employees acknowledging that it is imperative to bring reshare virality down on the platform, with one employee noting that “this change has the potential for high network effects since we are directly intervening into the reshare virality multipliers” – meaning that the fact that employees were deliberately reducing virality would have a significant impact on how far content travelled on the platform.²⁵⁸

From the Facebook Papers, it is clear that Meta was aware of the risks that its algorithmic systems continued to present, even with the introduction of the MSI metric. This knowledge means that Meta therefore had a responsibility to mitigate the harms of this feature, and that this responsibility was also heightened in a conflict-affected setting such as in Ethiopia, where risks were known to be most acute.

In November 2021, Meta published a blog post outlining some of the mitigation measures it was taking in Ethiopia, which included limiting the amount of content reshared by more than two people, in an acknowledgement of the risks this presents in a conflict-affected setting.²⁵⁹

However, as demonstrated by the case studies in Chapter 6, the announcement of this measure – one year after the outbreak of armed conflict – was too slow. Additionally, limiting reshares would not have prevented harmful content from going viral, particularly as the MSI metric amplifies content which receives a high number of comments and reactions, as well as reshares.

7.3 MSI PRIORITIZED OVER “INTEGRITY MEASURES”

Although the introduction of MSI supposedly heralded a new era of Meta prioritizing well-being to make sure time on Facebook was time “well-spent”,²⁶⁰ the Facebook Papers show that from 2018 until 2021, a year into conflict in northern Ethiopia, the company was still privileging the metric over platform changes which could reduce harms.

The Facebook Papers show a level of discomfort among some Meta employees with this approach. Discussing the new metrics that reduce the weighting given to reshares, one employee commented: “There have been many of us working to reduce reshares for the last year, but it’s hard to get something shipped²⁶¹ if it regresses MSI.”²⁶²

In a document discussing “soft actions” (an umbrella term for a wide range of options that Meta can apply that stop short of content removal) for content in crisis situations such as Ethiopia, there is evidence that suggests that Meta CEO Mark Zuckerberg intervened to stop the application of mitigation measures because of concerns around the impact on MSI – and, by extension, profitability. The author of the document writes: “Mark doesn’t think we should go broad... We wouldn’t launch if there were material trade-offs with MSI”.²⁶³ In this document, Ethiopia is mentioned as a country where experiments will be temporarily launched due to its status as an at-risk country.²⁶⁴

The privileging of engagement and the MSI metric over mitigation measures, particularly in a conflict-affected setting in Ethiopia is a clear dereliction of Meta’s responsibility under UNGP Principle 13 to respect human rights by avoiding causing or contributing to adverse human rights impacts through its activities, and crucially – to seek to address those impacts where they do occur.

²⁵⁷ The Facebook Papers, “MSI Metric changes for 2020 H1”, 22 January 2020, https://www.documentcloud.org/documents/21601827-tier2_rank_ro_0120, p. 1.

²⁵⁸ The Facebook Papers, “MSI Metric changes for 2020 H1” (previously cited), p. 15.

²⁵⁹ Meta, ‘An update on our longstanding work to protect people in Ethiopia’, 9 November 2021, (previously cited).

²⁶⁰ CNN Business, “Facebook to show more content from friends, less from publishers and brands” (previously cited).

²⁶¹ The phrase “ship” is often used in the tech industry to refer to the process of releasing or deploying a product.

²⁶² The Facebook Papers, “MSI Metric changes for 2020 H1” (previously cited), p. 10.

²⁶³ The Facebook Papers, “Discussion of ‘Soft Actions’ and Mark Zuckerberg’s feedback”, 20 April 2020, https://www.documentcloud.org/documents/23596616-tier3_force_ro_0420, p. 1.

²⁶⁴ The Facebook Papers, “Discussion of ‘Soft Actions’ and Mark Zuckerberg’s feedback” (previously cited), p. 1.

7.4 UNHEEDED WARNINGS

“I cried in the session and told them Ethiopia was going to be engulfed in war. I said there is going to be an insane amount of violence here. I was adamant that this was going to turn into something awful...”²⁶⁵

Hadush, Tigrayan community member

An important factor in assessing Meta’s responsibility for any abuses committed during the conflict in northern Ethiopia is the foreseeability of the company 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 the necessary steps to cease or prevent its contribution and use its leverage to mitigate any remaining impact to the greatest extent possible. Accordingly, this section outlines various warnings and interventions which Meta received between 2019 and 2022 in relation to Ethiopia.

Amnesty International has previously found that Meta’s content-shaping algorithms proactively amplified content which incited violence and hatred against the Rohingya, substantially contributing to the atrocities of 2017 in Myanmar.²⁶⁶ Amnesty International communicated these findings to Meta in 2022, giving the company an opportunity to respond.²⁶⁷ Meta did not respond to these specific findings.

Even before the outbreak of the conflict in northern Ethiopia, civil society organizations and human rights experts had warned that Meta did not have an adequate understanding of the complexities of Ethiopia’s socio-political context, which would be essential to effectively mitigate potential harms caused by the Facebook platform. Among others, the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, David Kaye, issued a call to Meta after meeting with civil society groups during his visit to Ethiopia in December 2019:

“[T]he Special Rapporteur urges Facebook and other information and communications technology companies to conduct periodic reviews of the human rights impact of their activities in Ethiopia, to establish more regular contact with the Government, relevant independent authorities and civil society to prevent or mitigate adverse impacts that may arise, and to consider opening branches in Ethiopia to better understand the issues at stake and ensure regular communication with the general public.”²⁶⁸

The Special Rapporteur also emphasized the need for better engagement with civil society groups monitoring digital platforms, writing: “As a first step, social media companies should establish regular and rapid-reaction mechanisms to enable civil society to report the most concerning kinds of content on these platforms.”²⁶⁹

Civil society organizations and human rights defenders told Amnesty International that they warned Meta about the impact the Facebook platform was having in Ethiopia, including the risk it could contribute to serious harms, at least since 2019 and in face-to-face meetings with the company.

Befekadu Hailu, the Executive Director of Centre for Advancement of Rights and Democracy (CARD), recounted a 2019 meeting which his organization facilitated between Meta’s East Africa team and 20 Ethiopian activists where the issue of algorithmic amplification of harmful content was raised:

“They [the activists] told them that Facebook shows them content they don’t like and the response from the official was that the algorithm doesn’t bring anything the user is not interested in... if you interact with that kind of content, you will see that kind of content... They said you determine the algorithm.”²⁷⁰

However, internal documents show that Meta employees were aware that simply because someone interacts with a piece of content, it does not mean that the content is not harmful to them or others. The company was also aware that there was an issue with its reporting mechanisms, which do not give users an option to flag content they interacted with but did not enjoy, or experienced as harmful. According to the Facebook

²⁶⁵ Amnesty International interview by video call with Hadush (pseudonym, real name withheld for security reasons), 26 April 2023.

²⁶⁶ Amnesty International, *The Social Atrocity: Meta and the right to remedy for the Rohingya* (previously cited).

²⁶⁷ Amnesty International letter to Meta, 13 June 2022.

²⁶⁸ UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Report: *Visit to Ethiopia*, 29 April 2020, UN Doc. A/HRC/44/49/Add.1, para. 38.

²⁶⁹ UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Report: *Visit to Ethiopia* (previously cited), para. 38.

²⁷⁰ Amnesty International interview by video call with Befekadu Hailu, Executive Director for the Centre for Advancement of Rights and Democracy, 24 July 2023.

Papers, one employee suggested: “I think we should directly explore a way for users to tell us that they didn’t like something and make it very easy to provide this feedback.”²⁷¹

Hadush, a Tigrayan activist, also attended a meeting with Meta’s East Africa team in 2019 to discuss social media and its impact in Ethiopia. Hadush told Amnesty International that he made it clear he believed that the Facebook platform was playing a significant role in the polarization of Ethiopian society:

“I said ‘this is going to be violent; this is going to be hateful; this is going to be horrific’. I told them I thought I was losing my nation because of it. It seemed like they hadn’t understood the impact of Facebook.”²⁷²

In July 2020, four months before the northern Ethiopia conflict began, a group of civil society organizations and human rights defenders wrote an open letter to Meta urging the company to take steps to ensure its services were not used to promote content which could incite violence and spread hate in Ethiopia.²⁷³ The letter warned that such content “can lead to physical violence and other acts of hostility and discrimination against minority groups.”²⁷⁴ The letter made several recommendations to prevent harms, including ceasing the algorithmic amplification of content inciting violence, temporary changes to sharing functionalities and conducting a human rights impact assessment into the company’s operations in Ethiopia.²⁷⁵

Meta responded to the letter three months later in October 2020, stating that the company had “researched, built and deployed multiple interventions to reduce the spread of inflammatory content in Ethiopia and improve integrity in other ways.”²⁷⁶

Meta runs a global “trusted partner programme”, which aims to build relationships with civil society organizations, to help the company better understand the impact of its operations and the context that its products, including the Facebook platform, operate in.²⁷⁷ Several of Meta’s “Trusted Partners” based in Ethiopia, who were asked to flag posts and accounts through a special reporting mechanism spoke out anonymously in the international news media about the complacency they faced from Meta staff when reporting problematic content on the platform – including the post which targeted Professor Meareg Amare.²⁷⁸ Trusted partners flagged the posts through a special channel and also in two Zoom meetings in October 2021. However, Meta agreed only to monitor the posts, rather than remove them completely. One anonymous trusted partner told Business Insider, who reported on the story: “One of the things Facebook said was, ‘We are not arbiters of the truth’. I remember asking, wouldn’t it be better if Facebook was taking down posts, [rather] than having posts stay on the platform that could hurt people.”²⁷⁹

Other “Trusted Partners” in Ethiopia reported that Meta was routinely slow to respond to urgent warnings, and that Meta staff would undermine their local expertise by disagreeing with their assessments of what should be considered a serious and actionable violation of Meta’s policies.²⁸⁰ Network Against Hate Speech, a civil society organization that became one of Meta’s trusted partners in January 2021, alleged that Meta chronically failed to reply to their messages and left material that had been flagged up on the Facebook platform for months.²⁸¹

Amnesty International spoke with Gelila, who works for an Ethiopian civil society organization which was involved in the Trusted Partner programme with Meta during the latter part of the conflict in northern Ethiopia, who highlighted many of the same issues. Gelila remarked particularly on Meta’s slow responses and their lack of respect for local knowledge:

²⁷¹ The Facebook Papers, “A proposal for bold experiments to learn about users and craft proxy metrics for Integrity” (previously cited), p. 6.

²⁷² Amnesty International interview by video call with Hadush, 26 April 2023.

²⁷³ Access Now, “Open letter to Facebook on violence – inciting speech: act now to protect Ethiopians”, 27 July 2020, <https://www.accessnow.org/open-letter-to-facebook-protect-ethiopians/>

²⁷⁴ Access Now, “Open letter to Facebook on violence – inciting speech: act now to protect Ethiopians” (previously cited).

²⁷⁵ Access Now, “Open letter to Facebook on violence – inciting speech: act now to protect Ethiopians” (previously cited).

²⁷⁶ Meta, “Ethiopia response”, 12 October 2020, <https://www.accessnow.org/wp-content/uploads/2020/12/Ethiopia-Response.pdf>

²⁷⁷ Meta, “Bringing local contexts to our global standards”, <https://transparency.fb.com/en-gb/policies/improving/bringing-local-context/>

²⁷⁸ Business Insider, “A professor was murdered after a Facebook post branded him a traitor. Was Facebook complicit in his death?”, 15 April 2023, <https://www.businessinsider.com/facebook-local-partners-say-hate-speech-stays-on-the-platform-2023-4?r=US&IR=T>

²⁷⁹ Business Insider, “A professor was murdered after a Facebook post branded him a traitor. Was Facebook complicit in his death?” (previously cited).

²⁸⁰ Business Insider, “A professor was murdered after a Facebook post branded him a traitor. Was Facebook complicit in his death?” (previously cited).

²⁸¹ Business Insider, “A professor was murdered after a Facebook post branded him a traitor. Was Facebook complicit in his death?” (previously cited).

“They are extremely slow in reacting to things. They are not sensitive to what is said – I think they have standards which are very far from what is happening on the ground. When you’re local you know what things trigger what, it might not work for their policy on whether it’s hate speech or not, they might have their own understanding in their office somewhere in the West, but on the ground you know what is hate speech in the local context.”²⁸²

The sluggish response of Meta to harmful content on the platform was echoed by Taye, the content moderator working in the Ethiopian market, who told Amnesty International:

“When you would mention things not being done – to do with the content moderators or the content itself – when we tried to give them feedback to update their policies, it was always very late. Late responses. Just the way the market was treated, it felt like it [the feedback process] was just a formality.”²⁸³

Hadush, a Tigrayan activist, was in touch with staff from Meta’s East Africa team intermittently throughout the conflict and reflected that the company took an inflexible approach to the situation in Ethiopia and the Facebook platform’s role:

“I didn’t feel like they had a system backing them [to] change how Meta worked to accommodate the situation...That wouldn’t be the case if Ethiopia was the US, the UK, France – the dynamism of the situation would be understood. They wanted to use already stringent existing rules about crisis to accommodate a completely different situation and a completely different environment.”²⁸⁴

Gelila also described warning Meta about the impact the Facebook platform was having in Ethiopia and the lack of adequate response:

“I told Facebook many times they were going to contribute to violence. Those people who are talking to you, they say they understand but it’s a big and bureaucratic organization. I don’t think they care much about what is happening on the ground.”²⁸⁵

During the northern Ethiopia conflict, Meta also received warnings about the risks it presented in the situation from its own Facebook Oversight Board. The Facebook Oversight Board was established in 2020 and is a panel of around 20 experts, including former political leaders, human rights activists and journalists picked by Meta to deliberate on the company’s decisions on content moderation.²⁸⁶ The Board takes cases which are brought either by Meta or by the public. While its decisions on content moderation are binding, its policy advisories are not.²⁸⁷

Prior to the outbreak of the conflict in northern Ethiopia, the Oversight Board had warned Meta about the impact its products may have in conflict-affected settings. In decision 2020-003-FB-UA, the Oversight Board found that “in situations of armed conflict in particular, the risk of hateful, dehumanizing expressions accumulating and spreading on the platform during an ongoing conflict, leading to offline action impacting the right of security to a person and potentially life is especially pronounced.”²⁸⁸

In an emblematic example of these problems, a Facebook user posted in July 2021 in Amharic, alleging that TPLF forces killed and raped women and children and looted properties in Raya Kobo and other towns in Ethiopia’s Amhara region. The Facebook user also alleged that Tigrayan civilians had assisted the TPLF in these atrocities, and that they (the poster) had received this information from people living in Raya Kobo. The user ended the post with the phrase “we will ensure our freedom through struggle”.²⁸⁹

Meta’s Amharic language classifier flagged the post and a human content moderator from the Amharic language team determined that the post violated Meta’s Hate Speech Community Standard and removed it. The user appealed the decision, which was confirmed by a second Amharic language content moderator.

²⁸² Amnesty International interview by video call with Gelila (pseudonym, name withheld for security reasons), 22 August 2023.

²⁸³ Amnesty International interview by video call with Taye, former Meta content moderator, 22 May 2023.

²⁸⁴ Amnesty International interview by video call with Hadush, 26 April 2023.

²⁸⁵ Amnesty International interview by video call with Gelila, 22 August 2023.

²⁸⁶ Facebook Oversight Board, <https://oversightboard.com/>; The New York Times, “What is the Facebook Oversight Board?”, 5 May 2021, <https://www.nytimes.com/2021/05/05/technology/What-Is-the-Facebook-Oversight-Board.html>

²⁸⁷ Facebook Oversight Board, <https://www.oversightboard.com/governance/>

²⁸⁸ Facebook Oversight Board, “Armenians in Azerbaijan”, <https://oversightboard.com/decision/FB-QBJDASCV/>

²⁸⁹ Facebook Oversight Board, “Alleged crimes in Raya Kobo”, <https://oversightboard.com/decision/FB-MP4ZC4CC/>

On 27 August 2021, Meta found that the decision to remove the post was incorrect and restored it to the Facebook platform.²⁹⁰ The post was then referred to the Oversight Board, which found that content violated Meta's Violence and Incitement Community Standard, which prohibits content containing language that incites or facilitates serious violence.²⁹¹

While it was available on the Facebook platform, the post was viewed 5,000 times.²⁹² Comments below the post included statements such as "our only option is to stand together for revenge" and "are you ready, brothers and sisters, to settle this matter?"²⁹³

In its policy advisory to Meta, issued in December 2021, the Oversight Board reflected on the centrality of the Facebook platform in Ethiopia citing expert briefings it had commissioned which demonstrated that "Facebook is an important, influential and popular online medium for communication in Ethiopia" and that "there is little to no coverage on the conflict-affected areas in Ethiopian media, and Ethiopians use Facebook to share and receive information about the conflict."²⁹⁴

The Oversight Board recommended that Meta should commission an independent human rights due diligence assessment on how Facebook and Instagram have been used to spread hatred and unverified rumours that heighten the risk of violence in Ethiopia.²⁹⁵ The Oversight Board proposed that the assessment should review the success of measures that Meta took to prevent the misuse of its products and services in Ethiopia as well as Meta's language capabilities in the country and whether they were adequate to protect the rights of platform users.²⁹⁶ The timeframe recommended for review was from 1 June 2020 (before the outbreak of the conflict in northern Ethiopia) to February 2021 and that it should take six months to complete.²⁹⁷

In January 2022, Meta responded that the company would "assess the feasibility" of conducting the independent human rights assessment recommended by the Board, claiming that such assessments can be "highly time intensive".²⁹⁸ As of October 2023, there is no evidence of any such human rights impact assessment on Ethiopia, or indeed if work on such an assessment has begun.

As demonstrated, Meta received repeated warnings from civil society groups, human rights defenders and "Trusted Partners" both before and throughout the conflict in northern Ethiopia, but continuously failed to take action which could have mitigated human rights impacts in Ethiopia. The corporate responsibility to respect human rights necessitates that companies track the effectiveness of their responses to adverse human rights impacts, including by drawing on the feedback of affected stakeholders. In dismissing the local knowledge of trusted partners and civil society activists, Meta failed to take into account valuable information which could have been used to mitigate – or even prevent – harms occurring on the ground in Ethiopia, including crimes under international law and other serious human rights violations.

In a letter to Amnesty International, Meta outlined its approach to stakeholder engagement in Ethiopia, stating that in 2022 members of the cross-functional team of Meta's staff visited Addis Ababa for three days to meet with a wide array of stakeholders including the Prime Minister's Office, various government agencies, human rights activists and civil society organisations.²⁹⁹ Meta staff undertook a similar visit in 2023.

In June 2021, before the Facebook Oversight Board recommendation, Meta finalized a human rights impact assessment on risk mitigation in the context of Ethiopia's elections and took some measures to mitigate harmful content, misinformation and out of context imagery.³⁰⁰ Meta outlined these steps in a blog post, but the full human rights impact assessment was not made publicly available. However, the severity of the risk outlined by the Facebook Oversight Board recommendation is such that Meta should have commissioned an impact assessment focused specifically on the spread of hatred and unverified rumours that heighten the risk of violence in Ethiopia.

As previously noted, transparency is a key component of human rights due diligence. The UN Guiding Principles make clear that companies need to "know and show that they respect human rights" and

²⁹⁰ Facebook Oversight Board, "Alleged crimes in Raya Kobo" (previously cited).

²⁹¹ Meta, Violence and incitement policy, <https://transparency.fb.com/en-gb/policies/community-standards/violence-incitement/>

²⁹² Facebook Oversight Board, "Alleged crimes in Raya Kobo" (previously cited).

²⁹³ Facebook Oversight Board, "Alleged crimes in Raya Kobo" (previously cited).

²⁹⁴ Facebook Oversight Board, "Alleged crimes in Raya Kobo" (previously cited).

²⁹⁵ Facebook Oversight Board, "Alleged crimes in Raya Kobo" (previously cited).

²⁹⁶ Facebook Oversight Board, "Alleged crimes in Raya Kobo" (previously cited).

²⁹⁷ Facebook Oversight Board, "Alleged crimes in Raya Kobo" (previously cited).

²⁹⁸ The Guardian, "Facebook owner to 'assess feasibility' of hate speech study in Ethiopia", 14 January 2022, <https://www.theguardian.com/technology/2022/jan/14/ethiopia-facebook-to-assess-feasibility-of-study-into-impact-on-human-rights>

²⁹⁹ Meta letter to Amnesty International, 18 September 2023

³⁰⁰ Meta, 'How Facebook is preparing for Ethiopia's 2021 general election', 10 June 2021, <https://about.fb.com/news/2021/06/how-facebook-is-preparing-for-ethiopias-2021-general-election/>

“showing involves communication, providing a measure of transparency and accountability to individuals or groups who may be impacted and to other relevant stakeholders.”³⁰¹

7.5 FAILURE TO ADEQUATELY MITIGATE RISKS

There are several pieces of evidence, primarily from the Facebook Papers and related media coverage, which show that Meta knew of the inadequacies of its mitigation measures in Ethiopia and the risks this presented in a country that the company itself considered to be at a priority and at high risk of violence. In this section, the extent of Meta’s internal knowledge and the quality of the mitigation measures which were taken will be considered.

The ability of Ethiopian platform users to report harmful content on the Facebook platform was an issue which was raised repeatedly in interviews Amnesty International conducted with affected individuals and digital rights experts, with many citing that the platform’s reporting mechanisms are unclear. Meta’s poor response time to reporting and its refusal to take down reported content also caused interviewees to feel that there was no point in reporting content.

In the case of Meareg Amare, Meta’s late response to reporting had tragic implications:

“I was trying to get attention from Facebook by reporting [the posts] several times in different subcategories. But I couldn’t get any response from Facebook until 11 November, which was eight days after the tragic loss of our father”³⁰²

Abrham Meareg, son of Professor Meareg Amare

Meta has been aware for several years that ensuring effective reporting processes was an issue in Ethiopia. An internal report from 2021 showed that even among at-risk countries, Ethiopia was an outlier with the lowest completion rate for user reports.³⁰³ The company attributed this to the lack of digital literacy in Ethiopia; however, the problem was compounded by its own reporting interfaces being confusing to platform users and a lack of local language reporting mechanisms.³⁰⁴

Other internal documents show that Meta was aware that the Facebook platform had been used to facilitate the spread of content advocating hatred in Ethiopia, and that its efforts to mitigate the harm were insufficient, as they have done little to mitigate the harms of algorithmically amplified content.³⁰⁵ An internal document from 2020 warned that current mitigation strategies are not enough to stop the spread of harmful content on the Facebook platform in Ethiopia.³⁰⁶ That same year, an internal audit reportedly found that Meta did not have automated detection systems for flagging messages advocating hatred in either Amharic or Oromo.³⁰⁷ Meta’s own researchers have repeatedly warned that, in general, the company appears ill-equipped to address issues such as content advocating hatred in languages other than English, potentially making users in some of the world’s most at-risk countries more vulnerable to physical violence.³⁰⁸

In a letter to Amnesty International dated 18 September 2023, Meta said that in 2021 they undertook work to improve and simplify reporting tools in Ethiopia, and that they have increased investment in active detection content moderation technology in the country.³⁰⁹ However, people interviewed by Amnesty International repeatedly raised poor response time and a general lack of understanding within Meta regarding harmful content on Facebook in Ethiopia, even after these improvements were made - strongly suggesting existing mitigation measures are insufficient.

³⁰¹ UN Guiding Principles, Commentary to Principle 21.

³⁰² Amnesty International interview by video call with Abrham Meareg, 11 April 2023.

³⁰³ Rest of World, “Why Facebook keeps failing in Ethiopia” (previously cited).

³⁰⁴ Rest of World, “Why Facebook keeps failing in Ethiopia” (previously cited).

³⁰⁵ The Atlantic, “How Facebook fails 90 percent of its users”, 25 October 2021, <https://www.theatlantic.com/ideas/archive/2021/10/facebook-failed-the-world/620479/>

³⁰⁶ CNN Business, “The big takeaways from the Facebook Papers”, 26 October 2021, <https://edition.cnn.com/2021/10/26/tech/facebook-papers-takeaways/index.html>

³⁰⁷ Wired, ‘A new lawsuit accuses Meta of inflaming civil war in Ethiopia’, 13 December 2022, <https://www.wired.com/story/meta-hate-speech-lawsuit-ethiopia/>

³⁰⁸ CNN Business, “Facebook has language blind spots around the world that allow hate speech to flourish” (previously cited).

³⁰⁹ Meta letter to Amnesty International, 18 September 2023

7.6 META'S KNOWLEDGE OF THE RISKS OF ALGORITHMIC AMPLIFICATION

From the Facebook Papers, it is clear that Meta has for many years been aware the risk that algorithmic amplification, coupled with a lack of mitigation measures, present to platform users. This section considers the impact of algorithmic amplification in the spread of content targeting the Tigrayan community on the Facebook platform.

Although Meta attempts to present its platform algorithms as content-agnostic, Amnesty International's 2022 *Social Atrocity* report's analysis of the Facebook Papers found evidence that employees had raised concerns about the use of algorithms optimized for engagement, since as early as 2016.³¹⁰ In one internal memo, an employee writes: "If I had to guess, the response to government regulation around engagement centric information feeds in 2026 will be 'Omg finally'."³¹¹

In an internal memo, dated 9 March 2020, a Meta employee stated: "Actively ranking content in News Feed and promoting content on recommendations surfaces makes us responsible for any harm caused by exposure to that content."³¹² The employee expands on the role that Meta's algorithms play in ranking content on Facebook, and how this is a choice which the company is responsible for:

"Facebook will generally have some responsibility for exposing users to harmful content whenever users saw that content as a result of Facebook's actions. Here, actions include anything Facebook does as an organization that influences the content users see – 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."³¹³

In a 2020 document, one Meta employee asks, "[A]re we also going to proactively look for issues in any of our engagement models that may be unduly incentivizing problematic content?"³¹⁴

Viral Facebook posts containing dehumanizing narratives, including messages advocating hatred that incited violence, hostility and discrimination against the Tigrayan community, were a dominant feature of the targeting of the community both inside and outside of the Tigray region during the armed conflict.

Abrehet Kahsay, a Tigrayan community member who was living in Bahir Dar, a city in the Amhara region of Ethiopia, at the beginning of the armed conflict, told Amnesty International about the impact of virality on a Facebook post which falsely accused her of being arrested for possessing weapons, posted by an anonymous account.

"[The post said] Abrehet Kahsay from BDU, she was caught carrying a bag full of hand bombs. We have to inspect all the Tigrayans living in our areas. And within a few minutes it had 300 likes... It was very dangerous and very scary to be honest."³¹⁵

These posts will necessarily have been ranked and spread by Meta's algorithms, as the algorithms determine what users see in their news feed and in what order, as well as what content is recommended to specific users.³¹⁶ As acknowledged by an internal employee in a memo from 2020, Meta should bear some level of responsibility for any harm which transpired as a result of the posts being viewed by a large number of people:

"When and whether a user sees a piece of content is... partly determined by the ranking scores our algorithms assign, which are ultimately under our control. This means, according to ethicists, Facebook is at least always partially responsible for any harmful experiences on News Feed."³¹⁷

³¹⁰ Amnesty International, *The Social Atrocity: Meta and the right to remedy for the Rohingya* (previously cited).

³¹¹ The Facebook Papers, "Why we build feeds", 4 October 2019, https://www.documentcloud.org/documents/21600853-tier1_rank_exp_1019, p. 1.

³¹² The Facebook Papers, "Facebook and Responsibility", 9 March 2020, https://www.documentcloud.org/documents/21594152-tier2_rank_other_0320, p. 1.

³¹³ The Facebook Papers, "Facebook and Responsibility" (previously cited), p. 3.

³¹⁴ The Facebook Papers, "Top-N Integrity", 23 September 2020, https://www.documentcloud.org/documents/21182838-tier2_el_ro_0920, p. 7.

³¹⁵ Amnesty International interview by video call with Abrehet Kahsay, 19 April 2023.

³¹⁶ Amnesty International, *The Social Atrocity: Meta and the right to remedy for the Rohingya* (previously cited).

³¹⁷ The Facebook Papers, "Facebook and Responsibility" (previously cited), p. 4.

Meta employees have also been aware that the content platform users see can have consequences beyond the digital domain. A 2019 internal memo notes that “[w]e have a pretty good idea that FB’s [Facebook’s] algorithms have a lot of unwanted side-effects.”³¹⁸

In an undated internal document discussing plans to increase the Facebook platform’s “defenses” against content which is harmful but does not violate company policy, a Meta employee admits that “[p]eople exposed to this content repeatedly may act in ways which are harmful to themselves, others, or society at large.”³¹⁹ The document highlights that at-risk countries are a priority for this plan, and specifically mentions Ethiopia.³²⁰

During her testimony to the United States Senate, whistle-blower Frances Haugen claimed that Meta’s algorithms “literally fan ethnic violence” in Ethiopia. Haugen elaborated:

“Facebook knows, they have admitted in public, that engagement-based ranking is dangerous without integrity and security systems, but then not rolled out those integrity and security systems to most languages in the world...And that’s what’s causing things like ethnic violence in places like Ethiopia”³²¹

Meta has been repeatedly criticized in media outlets and by digital rights experts for the unequal treatment given to English-speaking countries in the Global North and non-English speaking countries in the Global Majority in terms of mitigating the risks of algorithmic ranking systems.³²²

Meta has previously made changes to algorithmic ranking on the platform in crisis situations, as evidenced by documents from the Facebook Papers that show Meta employees experimented with additional mitigations in the US ahead of the 2020 election.,

A document from 23 September 2020 details an experiment internally dubbed “Top-N Integrity” designed with the November US election in mind. The new effort included daily reviews of the most popular content in Feeds, Stories, Pages, Rooms, Watch and Groups to identify “high-risk” content and investigate why the content was boosted. This included examining “distribution patterns (engagement and boosts), and if our integrity safeguards worked correctly.”³²³

Another Meta employee commented on the memo in Meta’s internal system Workplace, suggesting to implement this system in at-risk countries and noted: “Our integrity defenses are also *much* weaker in these markets, so integrity-minded ranking is particularly important here and would be high impact.”³²⁴ Ethiopia is specified as an at-risk country that would benefit from this change in algorithmic ranking.³²⁵

There is further evidence that employees at Meta were aware that algorithmic ranking and amplification needed greater mitigation measures in Ethiopia. In a memo discussing options for reducing inflammatory content, the author writes that the experiment which made “fundamental changes to the mechanisms of virality” was the most effective “against content that could potentially lead to violence in crisis situations like Sri Lanka and Ethiopia.”³²⁶

As noted above, in November 2021, Meta published a blog outlining mitigation measures it had taken in Ethiopia, which included limiting content reshared beyond two people.³²⁷ However, this measure would likely not have been sufficient to fully mitigate the risk of virality as a core feature in the Facebook platform, as content which received engagement such as comments and likes would still be prioritised for amplification by the platform’s algorithms.

³¹⁸ The Facebook Papers, “Integrity choice and LIKE data”, 13 November 2019, https://www.documentcloud.org/documents/23605720-tier2_civ_pr_1119, p. 5.

³¹⁹ The Facebook Papers, “Plan to Increase Facebook Defenses of Non-violating Harms”, undated, https://www.documentcloud.org/documents/23605717-tier2_civ_ro_undated, p. 1.

³²⁰ The Facebook Papers, “Plan to Increase Facebook Defenses of Non-violating Harms” (previously cited), p. 2.

³²¹ Rev, “Facebook whistleblower Frances Haugen testifies on children & social media use: Full Senate hearing transcript”, 5 October 2021, <https://www.rev.com/blog/transcripts/facebook-whistleblower-frances-haugen-testifies-on-children-social-media-use-full-senate-hearing-transcript>

³²² Mozilla Foundation, “Content moderation”, <https://foundation.mozilla.org/en/campaigns/trained-for-deception-how-artificial-intelligence-fuels-online-disinformation/content-moderation/>; Politico, “Facebook did little to moderate posts in the world’s most violent countries”, 25 October 2021, [https://www.politico.com/news/2021/10/25/facebook-moderate-posts-violent-countries-517050#:~:text=In%20many%20of%20the%20world's,and%20provided%20to%20Congress%20in](https://www.politico.com/news/2021/10/25/facebook-moderate-posts-violent-countries-517050#:~:text=In%20many%20of%20the%20world's,and%20provided%20to%20Congress%20in;); Rest of World, “The Facebook Papers reveal staggering failures in the Global South”, 26 October 2021, <https://restofworld.org/2021/facebook-papers-reveal-staggering-failures-in-global-south/>;

³²³ The Facebook Papers, “Top-N Integrity” (previously cited), p. 1.

³²⁴ The Facebook Papers, “Top-N Integrity” (previously cited), p. 7.

³²⁵ The Facebook Papers, “Top-N Integrity” (previously cited).

³²⁶ The Facebook Papers, “Big levers ranking experiment”, undated, https://www.documentcloud.org/documents/21602350-tier0_rank_ro_undated, p. 1.

³²⁷ Meta, “An Update on our longstanding work to protect the people of Ethiopia”, 9 November 2021, (previously cited)

In 2022, Amnesty International found that Meta had contributed to serious human rights violations against the Rohingya during the atrocities perpetrated by the Myanmar Military in Rakhine State in 2017.³²⁸ Amnesty International's analysis of Meta's failures to live up to its responsibility to respect human rights in Myanmar included the role that the Facebook platform's algorithms played in amplifying anti-Rohingya hatred and incitement to violence, a lack of Burmese-speaking content moderators, Meta's pre-existing knowledge of algorithmic harms and the risks this presented in Myanmar specifically and a pattern of ignoring warnings from civil society actors before and during the atrocities.³²⁹ Amnesty International shared these findings with Meta in a letter to the company in 2022. The company declined to the findings related to algorithmic amplification.

As outlined in the section above, three years after the 2017 atrocities against the Rohingya, Meta's surveillance-based business model once again fuelled the spread of harmful content in Ethiopia. The parallels between Myanmar and Ethiopia show that the risks Meta presents are systemic and will not improve until the company's business model is substantially changed or abandoned altogether.

7.7 META'S CONTRIBUTION TO HUMAN RIGHTS HARMS SUFFERED BY THE TIGRAYAN COMMUNITY

According to the UN Guiding Principles, if a business enterprise has contributed to an adverse human rights impact when its activities (including omissions) materially increase the risk of the specific impact which occurred – even if the business enterprise's activities would not have been sufficient in and of themselves to result in that impact.³³⁰ To fulfil its responsibility to respect human rights, Meta has a responsibility to “avoid 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.”³³¹

During the conflict in northern Ethiopia, the Facebook platform was used by a range of actors, such as the federal government, government-affiliated activists, and government-aligned news pages, to post content which targeted the Tigrayan community. Some of this content constituted advocacy of hatred inciting violence and discrimination. Meta's contribution to the negative human rights impacts suffered by the Tigrayan community stems from the role that the platform features – which constitute the foundation of its business model (particularly algorithmic amplification and, correspondingly, virality) – played in actively amplifying content which advocated hatred constituting incitement to violence, hostility and discrimination against the Tigrayan community. Additionally, the steps that the company took in 2020 and 2021 – such as translating its community standards into Amharic and Oromo, hiring more content moderators and limiting reshares – did not adequately address the algorithmic amplification of harmful content.

The effects of this were made more acute because the Facebook platform is Ethiopia's most widely used social media platform and is considered to be a valuable and trustworthy source of news and information in the country. Meta can also be considered to have contributed to adverse human rights impacts due to the foreseeability of the risk its operations presented in Ethiopia. Despite receiving repeated warnings by local and international civil society organizations, international human rights mechanisms, and its own Oversight Board, Meta failed to adequately mitigate the human rights risks of its operations in Ethiopia.

There are numerous additional steps that Meta could have taken to prevent the spread and amplification of content targeting the Tigrayan community and advocating hatred on the Facebook platform, such as more proactively implementing the suite of so-called “Break the Glass” measures, which have been designed specifically for crisis contexts. These measures are designed to reduce the power of Meta's algorithmic amplification.³³² The company has applied these measures in at least two cases, namely the 2020 US presidential elections and the January 2021 riots on Capitol Hill – both events happened within the same

³²⁸ Amnesty International, *The Social Atrocity: Meta and the right to remedy for the Rohingya* (previously cited).

³²⁹ Amnesty International, *The Social Atrocity: Meta and the right to remedy for the Rohingya* (previously cited).

³³⁰ Business & Human Rights Resource Centre, *Practical Definitions of Cause, Contribute, and Directly Linked to Inform Business Respect for Human Rights*, 9 February 2017, <https://www.business-humanrights.org/en/latest-news/practical-definitions-of-cause-contribute-and-directly-linked-to-inform-business-respect-for-human-rights/>

³³¹ 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” (previously cited), Principle 13 including Commentary.

³³² Meta, “Our Comprehensive Approach to Protecting the US 2020 Elections Through Inauguration Day”, 22 October 2021, <https://about.fb.com/news/2021/10/protecting-us-2020-elections-inauguration-day/>

timeframe of the conflict in northern Ethiopia.³³³ As they are designed specifically for crisis situations, implementing these same measures in a proactive manner in Ethiopia, among other mitigation interventions, could have mitigated, or prevented, the spread of content advocating hatred and the associated adverse human rights impacts, such as those outlined in the case studies in Chapter 6.

While some measures mitigating the risk of algorithmic ranking may have been applied in Ethiopia due to its status as a priority country for Meta, it is clear that more could have been done to adequately mitigate the risk of harm. In fact, Amnesty International has instead seen evidence suggesting that in April 2020, Mark Zuckerberg intervened to prevent the application of mitigation measures as they may have weakened the MSI engagement metric.³³⁴ Amnesty International sent a letter to Meta asking if the company had applied any mitigation, or ‘break the glass’ measures in Ethiopia, but the company did not provide a response to this question.³³⁵ As well as making changes to its algorithmic amplification and engagement metrics, Meta should have taken more robust action, much earlier, to improve its content moderation efforts. As detailed in Chapter 5, Meta was not able to adequately moderate content in the main languages spoken in Ethiopia and was slow to respond to feedback from content moderators regarding terms which should be considered harmful. This resulted in harmful content being allowed to circulate on the platform – at times even after it was reported, because it was not found to violate Meta’s community standards. While content moderation alone would not have prevented all the harms stemming from Meta’s algorithmic amplification, it is an important mitigation tactic.

Despite warnings from civil society – both before and during the armed conflict in northern Ethiopia – Meta did not make substantial changes to its engagement metrics or content-shaping algorithms in Ethiopia, continuing to amplify harmful content on the Facebook platform.

Amnesty International’s analysis of Meta’s role in serious human rights violations suffered by the Tigrayan community across Ethiopia between 2020 and 2022, based on international human rights standards including the UN Guiding Principles, leads to the following conclusions:

- At a time when Facebook was the dominant platform in Ethiopia, the federal government, government-affiliated activists and government-aligned news pages used Facebook to post content that targeted the Tigrayan community. Some of this content constituted advocacy of hatred inciting violence and discrimination.
- Meta’s failures of content moderation in Ethiopia allowed content which advocated hatred and incited violence against the Tigrayan community to spread on the Facebook platform.
- Meta’s content-shaping algorithms actively amplified and promoted divisive content, including messages advocating hatred, in the context of a country embroiled in an armed conflict.
- Meta knew or should have known that it risked contributing to human rights abuses in Ethiopia, having conducted multiple studies into the effects of its algorithms, its MSI metric, and receiving multiple warnings from its own Oversight Board, trusted partners, international human rights experts and local civil society activists with regards to risks in Ethiopia specifically.
- Meta failed to engage in adequate human rights due diligence, which could or should have identified the risks that its operations presented in Ethiopia. Meta also failed to enact adequate and appropriate mitigation measures which may have prevented or mitigated the harm in Ethiopia.
- In the case studies outlined in Chapter 6, Meta’s content-shaping algorithms, which boosted posts receiving a high number of comments and reactions, alongside the company’s failures of due diligence regarding the spread of content inciting violence in Ethiopia and its inadequate content moderation operations, contributed to the violations of a range of human rights, including the right to life, the right to security of person, and the right to freedom of expression.
- Meta contributed to adverse human rights impacts suffered by the Tigrayan community, and therefore has a corresponding responsibility to remediate the harm experienced by affected individuals.

³³³ Meta, “Our Comprehensive Approach to Protecting the US 2020 Elections Through Inauguration Day” (previously cited); Politico, “Inside Facebook’s struggle to contain insurrectionists’ posts”, 25 October 2021, <https://www.politico.com/news/2021/10/25/facebook-jan-6-election-claims-516997>

³³⁴ The Facebook Papers, “Discussion of ‘Soft Actions’ and Mark Zuckerberg’s feedback” (previously cited).

³³⁵ Amnesty International letter to Meta, 18 July 2023

8. MAKE IT RIGHT: REMEDY AND PREVENTION OF FUTURE HARM

“No one needs to suffer as my family has, ever again. We don’t want to allow anyone else to suffer. And we are stood in the court, in front of justice, on behalf of other victims who cannot speak out.”³³⁶

Abrham Meareg

As outlined in the chapter above, Meta contributed to serious human rights harms suffered by the Tigrayan community in the context of the 2020 – 2022 armed conflict in northern Ethiopia. As a result, in accordance with the UN Guiding Principles, the company has a responsibility to provide effective remedies and adequate reparations to those who have been affected by their operations in the context of the armed conflict in Ethiopia. This chapter outlines Meta’s responsibility to provide remedy for the harms it has contributed to in Ethiopia.

Under international law, the appropriate type of remediation depends on the nature of the harm, but it should include compensation, rehabilitation, measures of satisfaction and guarantees of non-repetition. This chapter will focus on ongoing processes to provide remedy and reparations to the victims in Ethiopia and the need to ensure that measures of non-repetition are put in place as tensions and violence continue to foment across Ethiopia and in other conflict-affected settings across the globe.

8.1 JUSTICE AND ACCOUNTABILITY: CIVIL LITIGATION IN KENYA

For Abrham Meareg and Fisseha Tekle, two of the petitioners in the civil litigation filed in December 2022 against Meta in Kenya pertaining to the Facebook platform’s impacts in Ethiopia, the pursuit of justice and remedy is both a matter of principle and an urgent need to ensure that similar harms are not repeated

³³⁶ Amnesty International interview by video call with Abrham Meareg, 11 April 2023.

elsewhere. In April 2023, the Kenyan court granted the petitioners leave to serve court papers to Meta in the USA.³³⁷

The legal case – to which Amnesty International is an interested party³³⁸ – argues that Meta promoted content that led to ethnic violence, killings, and other crimes under international law in Ethiopia by utilizing an algorithm that prioritizes and recommends content advocating hatred that incites violence, hostility, and discrimination.³³⁹

Additionally, the legal action alleges the Facebook platform in Ethiopia was – and continues to be – awash with hateful, inciteful and harmful posts, and that the posts which targeted Meareg Amare and Fisseha Tekle are not isolated cases.³⁴⁰ Abraham Meareg lamented the lack of accountability Meta has faced for its contribution to human rights impacts in Ethiopia:

“I am very disappointed; it is disappointing that they still continue as if nothing happened. As if things are normal and as if the safety of their platform is effective. But it’s quite the opposite.”³⁴¹

Fisseha Tekle explained to Amnesty International why he believes that it is important that Meta acknowledges the harm caused by providing remedy:

“Compensation and rehabilitation won’t give it all back but it’s symbolic, acknowledges their contribution and shows goodwill. As part of the remedy, they also have to look at their business practices, their algorithms and content moderation and all these things. That’s very important.”³⁴²

Although she is not involved in the civil litigation against Meta, Frewyni Itsay – who lost her father after the circulation of Facebook posts inciting violence against him – explained to Amnesty International why she strongly feels that Meta must be answerable for the human rights impacts it contributed to in Ethiopia:

“Lives were lost, infrastructure was destroyed, many things were destroyed. And Facebook has a responsibility to make things right for the coming generation.”³⁴³

Meta has publicly responded to the litigation saying that hate speech and incitement to violence are against the rules of the Facebook platform, and that it invests heavily in teams and technology to help them find and remove violating content.³⁴⁴

8.2 META’S RESPONSIBILITY TO PROVIDE REMEDY

Companies that have contributed to adverse human rights impacts have a responsibility to adequately remediate those affected.³⁴⁵ The appropriate type of remediation depends on the nature of the harm and may take a range of forms, including apologies, restitution, rehabilitation, financial or non-financial compensation, and justice (both through criminal or administrative mechanisms), as well as guarantees of non-repetition for the prevention of future harm.³⁴⁶

³³⁷ TechCrunch, “Kenya court paves way for lawsuit alleging Facebook played role in fueling Ethiopia’s Tigray conflict”, 28 April 2023, <https://techcrunch.com/2023/04/28/kenyan-court-paves-way-for-lawsuit-alleging-facebook-played-role-in-fuelling-ethiopias-tigray-conflict/>

³³⁸ For a definition of interested party, see: Kenya Law, *Land Case No.27 of 2021*, <http://kenyalaw.org/caselaw/cases/view/221027#:~:text=Rule%20of%20the%20Mutunga,person%20to%20mean%20both%20a,para.14.>

³³⁹ Amnesty International, “Kenya: Meta sued for 1.6 billion USD for fueling Ethiopia ethnic violence” (previously cited).

³⁴⁰ Amnesty International, “Kenya: Meta sued for 1.6 billion USD for fueling Ethiopia ethnic violence” (previously cited).; Kenyan Constitutional Court, *2nd Petitioner’s supporting affidavit to the petition*, December 2022, <https://www.foxglove.org.uk/wp-content/uploads/2022/12/2nd-Petitioners-Supporting-Affidavit.pdf>, paras. 122-155.

³⁴¹ Amnesty International interview by video call with Abraham Meareg, 11 April 2023.

³⁴² Amnesty International interview by video call with Fisseha Tekle, 19 April 2023.

³⁴³ Amnesty International interview by video call with Frewyni Itsay, 12 May 2023.

³⁴⁴ Reuters, “Meta accused in lawsuit of allowing posts that inflamed Ethiopia conflict”, 15 December 2022, <https://www.reuters.com/world/africa/lawsuit-accuses-meta-enabling-hateful-posts-ethiopia-conflict-2022-12-14/>

³⁴⁵ OECD Due Diligence Guidelines; UN Guiding Principle 22.

³⁴⁶ UN Guiding Principles Interpretive Guide, p. 7.

A public apology is an important form of remediation, which acknowledges the facts and acceptance of responsibility, which could be accompanied by verification of the facts and full and public disclosure of the truth.³⁴⁷

An equally important form of remediation is a guarantee of non-repetition, which is intended to prevent abuses from occurring in future. In this context, the prevention of further abuses can be achieved through several measures including both regulatory and accountability measures to be taken by states, and actions to be taken by companies themselves – any or all of which could contribute to guaranteeing non-repetition in the future.³⁴⁸ Among other things, efforts to guarantee non-recurrence could include the publication of the human rights impact assessment into Meta’s operations in Ethiopia, and reforming Meta’s content-shaping algorithms.

The civil litigation case in Kenya explicitly requests a change to Meta’s business model as a guarantee of non-repetition to be considered part of the remedy for Abraham Meareg and Fisseha Tekle. In Meta’s case this aspect of its responsibility to provide effective remedy is crucial not only for Abraham Meareg and Fisseha Tekle, but for the human rights of at-risk communities across the world. The urgency of this issue is more pronounced given that Meta has previously been found to have contributed to serious human rights abuses in a conflict-affected setting – and because of the continued risk of conflict in Ethiopia.³⁴⁹

8.3 RISK OF FUTURE HARM IN ETHIOPIA

8.3.1 ETHIOPIA’S TELECOMS EXPANSION

In 2019, Ethiopia’s Prime Minister declared that Ethiopia would privatize and liberalize its economy to encourage competition in several critical sectors, including telecommunications, reversing decades of policy by opening Ethiopia’s telecoms state monopoly to foreign competition.³⁵⁰ It is expected that this initiative will lead to the expansion of broadband access in the country.³⁵¹

In 2020, the framework for Ethiopia’s telecoms transformation was launched with the government’s “Digital Ethiopia 2025”, with network expansion and mobile connectivity designated as priority projects.³⁵² The process was postponed in March 2022 due to concerns that the country would not be able to attract an adequate amount of foreign investment³⁵³ but was restarted by November of that year.³⁵⁴ Commentators have noted that Ethiopia’s liberalization may be the most significant since Myanmar’s in 2013 and has the potential to revolutionize the country’s telecoms sector.³⁵⁵

The liberalization of Myanmar’s underdeveloped telecoms sector could serve as a useful predictor of the impact that similar changes will have in Ethiopia. Myanmar’s telecoms liberalization led to an explosion in mobile and internet penetration – in 2011, mobile penetration in Myanmar was estimated at 2%, while internet penetration was at just 0.23%.³⁵⁶ By 2017, mobile penetration had reached 93% and internet penetration was at 26%.³⁵⁷ Amnesty International has previously found that the rapid expansion of internet penetration in Myanmar, combined with Meta’s algorithmic amplification of content advocating hatred inciting to violence and failures to undertake adequate due diligence, resulted in the Facebook platform

³⁴⁷ 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.

³⁴⁸ Amnesty International, *Injustice Incorporated: Corporate abuses and the human right to remedy* (Index: POL 30/001/2014), 7 March 2014, <https://www.amnesty.org/en/documents/pol30/001/2014/en/>, p. 18.

³⁴⁹ Amnesty International, *The Social Atrocity: Meta and the right to remedy for the Rohingya* (previously cited).

³⁵⁰ Financial Times, “My model is capitalism: Ethiopia’s prime minister plans telecoms privatization”, 24 February 2019, <https://www.ft.com/content/433dfa88-36d0-11e9-bb0c-42459962a812>

³⁵¹ Federal Democratic Republic of Ethiopia, “Digital Ethiopia 2025 – A strategy for Ethiopia inclusive prosperity”, https://www.lawethiopia.com/images/Policy_documents/Digital-Ethiopia-2025-Strategy-english.pdf

³⁵² Federal Democratic Republic of Ethiopia, “Digital Ethiopia 2025 – A strategy for Ethiopia inclusive prosperity” (previously cited).

³⁵³ Ethiopia’s Ministry of Finance, Twitter post, 18 March 2022, https://twitter.com/MoF_Ethiopia/status/1504785875783331872?ref_src=twsrc%5Egoogle%7Ctwcamp%5Eserp%7Ctwgr%5Etweet

³⁵⁴ Reuters, “Ethiopia restarts sale of Ethio Telecom stake, new telecoms licence”, 16 November 2022,

<https://www.reuters.com/markets/deals/ethiopia-revives-sale-ethio-telecom-stake-new-telecoms-licence-2022-11-16/>

³⁵⁵ Financial Times, “Ethiopia’s promising but problematic telecoms privatization”, 7 June 2021, <https://www.ft.com/content/deabd4ae-4bc0-4b41-bb25-cebec2ddc560>

³⁵⁶ Datareportal, “Digital 2011: Myanmar”, 28 December 2011, <https://datareportal.com/reports/digital-2011-myanmar>

³⁵⁷ DataReportal, “Digital 2017: Myanmar”, 1 February 2017, <https://datareportal.com/reports/digital-2017-myanmar>

contributing to crimes under international law and other serious human rights violations against the Rohingya in 2017.³⁵⁸

Increased internet access is crucial in the modern world and will bring many benefits to people in Ethiopia. However, in the absence of effective human rights due diligence and adequate mitigation measures by Meta, it may also result in the Facebook platform presenting an even more significant risk to the human rights of platform users across the country.

Meta has already contributed to serious human rights abuses due to a combination of the surveillance-based business model, content-shaping algorithms, a lack of adequate resourcing for the Ethiopian market, and failure to conduct adequate human rights due diligence. Over the next two years, it is expected that this market will rapidly expand, with more people gaining access to the Facebook platform across Ethiopia. It is imperative that Meta considers the risks that its operations present in Ethiopia before this market expansion and takes meaningful actions to adequately prevent and mitigate those risks.

It should be noted that despite plans to broaden internet access in Ethiopia, the government has frequently shut down access to the internet in times of crisis. During the conflict in northern Ethiopia, the Tigray region was subject to a telecommunications shutdown for nearly two years.³⁵⁹ In February 2023, access to social media platforms, including Facebook, was blocked following a disagreement in Ethiopia's Orthodox Church, which saw church leaders threaten to call for country wide rallies and counter rallies.³⁶⁰ These measures are disproportionate, and also violate the right to freedom of expression, which includes the freedom to seek, receive and impart information.³⁶¹ Therefore, limiting access to the Facebook platform should not be considered an appropriate response to the risks the platform presents in Ethiopia.

8.3.2 RISK OF FUTURE ATROCITIES IN ETHIOPIA

On 2 November 2022, the Ethiopian government and the TPLF formally agreed to a permanent cessation of hostilities.³⁶² However, there remain concerns that human rights violations are continuing to occur in northern Ethiopia and elsewhere in the country. Additionally, Amnesty International has found that crimes under international law took place during the final days of the negotiations for the cessation of hostilities agreement (CoHA), between 25 October and 1 November 2022, with the Eritrean Defense Forces (EDF) extrajudicially executing civilians in Mariam Shewito, a district in the Tigray region.³⁶³ Furthermore, the risk of violence in Ethiopia remains high. According to the UN's 2014 Atrocity Prevention Framework, states where atrocities have already occurred are more likely to have them occur again.³⁶⁴

The Framework outlines 14 risk factors for the occurrence of atrocity crimes.³⁶⁵ Risk factors identified in the framework include "behaviours, circumstances or elements that create an environment conducive to the commission of atrocity crimes, or indicate the probability".³⁶⁶ These include eight "common risk factors", which are designed to identify the probability of atrocity crimes overall, without necessarily identifying the type of crime.³⁶⁷

There are common risk factors listed in the Framework which Meta may contribute to, unless it takes adequate steps to mitigate the risks presented by the Facebook platform, in particular:

³⁵⁸ Amnesty International, *The Social Atrocity: Meta and the right to remedy for the Rohingya* (previously cited).

³⁵⁹ Access Now, "Two years of internet shutdowns: people in Tigray, Ethiopia, deserve better" (previously cited).

³⁶⁰ Amnesty International, "Ethiopia: One month on, authorities must immediately lift blockade on selected social media access in the country", 9 March 2023, <https://www.amnesty.org/en/latest/news/2023/03/ethiopians-in-social-media-blackout-for-second-month/>

³⁶¹ UNHCHR, "Internet shutdowns: trends, causes, legal implications and impacts on a range of human rights" (previously cited); Amnesty International, "Ethiopia: One month on, authorities must immediately lift blockade on selected social media access in the country" (previously cited).

³⁶² AP News, "Ethiopian govt, Tigray agree to end fighting after 2 years", 2 November 2022, <https://apnews.com/article/africa-south-ethiopia-african-union-70fb0d185aaccb668b2fabb4f6e45b9f>

³⁶³ Amnesty International, *Today or tomorrow, they should be brought before justice: Rape, sexual slavery, extrajudicial executions, and pillage by Eritrean forces in Tigray* (Index: AFR 25/7152/2023), 4 September 2023, <https://www.amnesty.org/en/documents/afr25/7152/2023/en/>

³⁶⁴ United Nations Office on Genocide Prevention and the Responsibility to Protect, "Framework of analysis for atrocity crimes: A tool for prevention", 30 October 2014, <https://www.globalr2p.org/resources/framework-of-analysis-for-atrocity-crimes-a-tool-for-prevention/#:~:text=During%20October%202014%20the%20UN,act%20early%2D%20to%20prevent%20atrocities.>

³⁶⁵ United Nations Office on Genocide Prevention and the Responsibility to Protect, "Framework of analysis for atrocity crimes: A tool for prevention" (previously cited).

³⁶⁶ United Nations Office on Genocide Prevention and the Responsibility to Protect, "Framework of analysis for atrocity crimes: A tool for prevention" (previously cited), p. 5.

³⁶⁷ United Nations Office on Genocide Prevention and the Responsibility to Protect, "Framework of analysis for atrocity crimes: A tool for prevention" (previously cited), p. 5.

- Common Risk Factor 4: Motives and Incentives – according to the Framework, “from an early warning perspective, it is extremely important to be able to identify motivations, aims or drivers that could influence certain individuals or groups to resort to massive violence as a way to achieve goals, feed an ideology and respond to real or perceived threats... *The historical, political, economic or even cultural environment in which such ideologies develop can also be relevant*”³⁶⁸ (emphasis added).
- Common Risk Factor 7: Enabling circumstances or preparatory action. Indicator 7.14 – Increased inflammatory rhetoric, propaganda campaigns or hate speech targeting protected groups, populations or individuals. Echoing the academic research on genocide and mass violence discussed previously in this report, the Framework notes that this risk factor can “create an environment that favours or even encourages the commission of [atrocities] crimes.”³⁶⁹
- Common Risk Factor 8: Triggering factors. Indicator 8.7 – Acts of incitement or hate propaganda targeting particular individuals or groups. The Framework states the importance of taking all possible triggering factors into account: “An adequate early warning assessment should be mindful of all such events and circumstances and consider their *potential impact, even if they appear unrelated to more direct or structural risk factors*”.³⁷⁰

On 4 August 2023, the Ethiopian government declared a six-month state of emergency following increased violence in the Amhara region.³⁷¹ The state of emergency, which was approved by Ethiopia’s House of People’s Representatives on 14 August 2023, gives the government sweeping powers to arrest people without a court warrant, impose curfews, prevent freedom of movement, and ban public assemblies or associations.³⁷²

In the same month, reports of gross human rights violations emerged following armed confrontations between the Ethiopian National Defense Force (ENDF) and local armed Fano militia.³⁷³ Amnesty International has also received allegations of mass killings and casualties in Finote Selam, Bahir Dar and Shewa Robit which merit further investigation.³⁷⁴ The Ethiopian government has also announced that it has detained individuals in Addis Ababa in informal detention centres, including schools.³⁷⁵

The state of emergency also allows the closure or restriction of media outlets or other entities suspected of “acting contrary to the purposes of this proclamation.” Not only are these terms ill-defined and over-broad, but the suppression of accurate media reporting on the political situation of the country may result in the Facebook platform becoming an even more important source of information for people in Ethiopia than it already is.

Gelila, who works for one of Meta’s former “Trusted Partners”, outlined her fears about the impact the Facebook platform could have in this new conflict and in the context of a potential explosion of new users as Ethiopia’s internet connectivity improves:

“Even when a small number of the community had Facebook, it was used to mobilize a significant number in Ethiopia to protest against the government. Now it is being used to mobilize against each other, between Amhara and Tigray and Amhara and Oromo. For me there will be major, major violence because of social media.”³⁷⁶

It is therefore urgent that Meta takes steps to mitigate the human rights impacts of its operations in Ethiopia to avoid continuing to contribute to serious abuses in the country.

³⁶⁸ United Nations Office on Genocide Prevention and the Responsibility to Protect, “Framework of analysis for atrocity crimes: A tool for prevention” (previously cited), p. 13

³⁶⁹ United Nations Office on Genocide Prevention and the Responsibility to Protect, “Framework of analysis for atrocity crimes: A tool for prevention” (previously cited), p. 16.

³⁷⁰ United Nations Office on Genocide Prevention and the Responsibility to Protect, “Framework of analysis for atrocity crimes: A tool for prevention” (previously cited), p. 17.

³⁷¹ The Guardian, “Ethiopia declares a state of emergency in Amhara amid increasing violence”, 4 August 2023, <https://www.theguardian.com/global-development/2023/aug/04/ethiopia-declares-a-state-of-emergency-in-amhara-amid-increasing-violence>

³⁷² Amnesty International, “Ethiopia: Authorities must grant independent investigators, media unfettered access to Amhara region to probe violations under state of emergency”, 18 August 2023, <https://www.amnesty.org/en/latest/news/2023/08/ethiopia-authorities-must-grant-independent-investigators-media-unfettered-access-to-amhara-region-to-probe-violations-under-state-of-emergency/>

³⁷³ Amnesty International, “Ethiopia: Authorities must grant independent investigators, media unfettered access to Amhara region to probe violations under state of emergency” (previously cited).

³⁷⁴ Amnesty International, “Ethiopia: Authorities must grant independent investigators, media unfettered access to Amhara region to probe violations under state of emergency” (previously cited).

³⁷⁵ Amnesty International, “Ethiopia: Authorities must grant independent investigators, media unfettered access to Amhara region to probe violations under state of emergency” (previously cited).

³⁷⁶ Amnesty International interview by video call with Gelila, 22 August 2023.

9. CONCLUSION AND RECOMMENDATIONS

“It’s the same story over and over again. Nothing changes. In Myanmar, people said the same thing. The same thing will happen somewhere else”.³⁷⁷

Timnit Gebru

CONCLUSION

This report, based on a thorough investigation of Meta’s role in serious human rights abuses perpetrated against the Tigrayan community, has firmly established that the company contributed to these harms and, therefore, has a corresponding responsibility to provide remedy to affected individuals and to urgently take additional mitigation measures to prevent future harms both in Ethiopia and in other conflict-affected settings worldwide.

As a global company that operates in high-risk and conflict-affected settings in every region of the world, there is a major and current risk that Meta’s operations could fuel advocacy of hatred and incite violence against ethnic and religious minorities across many other parts of the world. The alarm has already been raised in multiple contexts and over many years. Whistle-blower Frances Haugen repeatedly warned that Meta is repeating its failures in Myanmar in other countries – including in Ethiopia, saying:

“What we saw in Myanmar and are seeing now in Ethiopia are only the beginning chapters of a story so terrifying no one wants to read to the end of it”.³⁷⁸

Haugen specifically highlighted Meta’s content-shaping algorithms as the key driver of these risks and harms.³⁷⁹ In other countries, such as India and Sri Lanka, the spread of advocacy of hatred that incites violence, hostility and discrimination has been linked to the Facebook platform’s content-shaping algorithms.³⁸⁰

In a context where social polarization had been present for decades, and with escalating dehumanizing political rhetoric against the Tigrayan community being used, Meta’s surveillance-based business model and engagement-centric algorithms helped to normalize hate, violence and discrimination against the Tigrayan

³⁷⁷ Amnesty International interview by video call with Timnit Gebru, 7 July 2023.

³⁷⁸ Rev, “Facebook whistleblower Frances Haugen testifies on children & social media use: Full Senate hearing transcript” (previously cited).

³⁷⁹ The Guardian, “Facebook’s role in Myanmar and Ethiopia under new scrutiny”, 7 October 2021, <https://www.theguardian.com/technology/2021/oct/07/facebooks-role-in-myanmar-and-ethiopia-under-new-scrutiny>

³⁸⁰ Time, “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, <https://time.com/5712366/facebook-hate-speech-violence/>; Article One, “Assessing the human rights impact of the Facebook platform in Sri Lanka”, 2018, <https://about.fb.com/wp-content/uploads/2020/05/Sri-Lanka-HRIA-Executive-Summary-v82.pdf>

community. The company's engagement-centric algorithms actively amplified and promoted harmful content, and the MSI metric encouraged comments and user interactions on this content, contributing to the spread of incitement to violence and the steady dehumanization of the Tigrayan community. At the same time, the company consistently failed to act upon repeated warnings it received from civil society – and from its own employees – about the risk it could contribute to mass violence in Ethiopia.

Meta's failures in Ethiopia had a similar effect to its recklessness in Myanmar: it poured fuel on the fire of already existing ethnic tensions and ultimately contributed to serious human rights abuses suffered by the Tigrayan community.

In 2022, Amnesty International similarly found that Meta had contributed to the commission of serious human rights violations perpetrated against the Rohingya during the atrocities by the Myanmar Military in Rakhine State in 2017.³⁸¹ Amnesty International's analysis of Meta's failures to live up to its responsibility to respect human rights in Myanmar included the role that the Facebook platform's algorithms played in amplifying anti-Rohingya hatred and incitement to violence, a lack of Burmese-speaking content moderators, Meta's pre-existing knowledge of the risk of algorithmic harms, and a pattern of ignoring warnings from civil society actors before and during the atrocities.³⁸²

The parallels between the experience of the Rohingya and the experience of Ethiopian Facebook users was commented on by journalist Zecharias Zelalem:

“I don't think anyone has had it worse than the Rohingya and the Ethiopians, I think the two will be bound together for future researchers to see how horrible things get with social media.”³⁸³

The fact that Meta has once again contributed to severe human rights abuses – in a conflict which began three years after the atrocities against the Rohingya in 2017 – raises serious questions about the adequacy of its human rights due diligence and its willingness to take appropriate and effective mitigation measures, particularly in conflict-affected settings. The company's apparent failure to conduct and publish a human rights impact assessment focused on the spread of hate and the connection to offline violence in Ethiopia – ignoring the recommendations of its own Oversight Board – is especially concerning and represents yet another attempt to avoid transparency and accountability.

Although Meta took several mitigation measures, such as improving its content moderation and language classifier systems in Ethiopia and to reduce the number of reshares, over the course of the conflict in northern Ethiopia, these reforms were too late for many victims and significantly below the level required to adequately mitigate the negative human rights impact of its operations. Moreover, these measures are too limited in scope and will not be able to provide a guarantee of non-repetition as they do not address the root cause of the threat Meta represents to human rights – the company's data-hungry business model.

The findings of this research should make clear how urgently changes are needed to Meta's business model: there remains a serious risk that the company will continue to contribute to serious human rights abuses in Ethiopia and other conflict-affected settings across the world. There can be no doubt that Meta's algorithms are capable of harming societies across the world by promoting content that advocates hatred and which incites violence and discrimination, which disproportionately impacts already marginalised communities.

Despite multiple warnings from civil society and all the reassurances given by Meta, the company has repeated many of its prior failings in Ethiopia. The fact that Meta's response to the conflict in Ethiopia still came up short even after its failures during the conflict in Myanmar is the most damning evidence available that the company will not solve the problems presented by its business model of its own accord. It has once again shown that it is either unwilling or incapable of addressing these issues in the absence of robust regulation. It should not be allowed to make the same mistakes for a third time.

Meta's lack of acknowledgement or accountability for the harms caused by its business operations in Ethiopia demonstrate more clearly than ever the company is incapable or unwilling to make the much-needed improvements to sufficiently and efficiently mitigate the harms associated with its content-shaping algorithms which are at the crux of its business model. Meta should urgently change course and seek to improve its operations to uphold its responsibility to respect human rights, including through adequate resourcing of content moderation and building more meaningful relationships with civil society organizations in Ethiopia and in all the countries in which it operates.

³⁸¹ Amnesty International, *The Social Atrocity: Meta and the right to remedy for the Rohingya* (previously cited).

³⁸² Amnesty International, *The Social Atrocity: Meta and the right to remedy for the Rohingya* (previously cited).

³⁸³ Amnesty International interview by video call with Zecharias Zelalem, 20 April 2023.

The unregulated development of Big Tech has resulted in grave human rights consequences around the world. A central piece 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. It is more crucial than ever that states honour their obligation to protect human rights by introducing and enforcing meaningful legislation that will rein in the surveillance-based business model.

RECOMMENDATIONS

TO META

REMEDY AND PREVENTION OF FUTURE HARM IN ETHIOPIA

- Publicly acknowledge and apologise for Meta’s contributions to human rights harms in Ethiopia, including by issuing a direct apology to Abraham Meareg, Fisseha Tekle and other individuals that were directly affected by the spread of advocacy of hatred on Facebook’s platform.
- Commit to fundamentally changing Meta’s surveillance-based business model to ensure these harms are not repeated again in the future.
- Establish a restitution fund for affected victims of the conflict in northern Ethiopia – and groups from other conflict-affected settings globally.
- Reform the “Trusted Partner” programme in Ethiopia, allowing civil society organizations and human rights defenders to have more meaningful input into content-related decisions.
- Expand Meta’s capacity to moderate in all 84 languages used in Ethiopia, including by directly hiring more content moderators and in working conditions which adhere to human rights standards.

HUMAN RIGHTS DUE DILIGENCE

- Conduct and publish a comprehensive human rights impact assessment on Ethiopia as soon as possible, as recommended by the Facebook Oversight Board.
- Publish and regularly update the list of countries internally identified as being “at risk”, alongside the mitigation measures being taken in each country to ensure that Meta does not contribute to serious human rights abuses.
- Conduct human rights impact assessments for all countries considered to be “at risk” and make the findings public.
- Ensure that the human rights due diligence processes used at Meta address the human rights impacts of Meta’s business model as a whole and prioritize transparency in how risks are identified and addressed.
- Ensure that human rights impact assessments are conducted in relation to the design and deployment of new AI systems and engagement metrics, such as Meaningful Social Interactions (MSI), and ensure meaningful public consultation.
- Undertake constant, ongoing and proactive human rights due diligence throughout the lifecycle of algorithmic technologies, including 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 technologies have been launched.

BUSINESS MODEL AND ALGORITHMS

- Cease the invasive collection of personal data which undermines the right to privacy and threatens a range of human rights.
- End the practice of using targeted advertising and embrace a less harmful alternative business model.
- Provide an opt-in for the use of content-shaping algorithms, so platform users have more control over their online experience.

“A DEATH SENTENCE FOR MY FATHER”

META’S CONTRIBUTION TO HUMAN RIGHTS ABUSES IN NORTHERN ETHIOPIA

Amnesty International

- Introduce “friction” measures as the norm, rolling out measures which have proven to be effective at improving “integrity” outcomes in crisis situations as standard, such as limits on resharing, message forwarding, and group sizes.
- Implement “break the glass” measures in high-risk situations proactively and as soon as they arise.
- 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 data, which is in the public interest, including data pertaining to algorithmic systems.

REGARDING BUSINESS OPERATIONS IN GLOBAL MAJORITY COUNTRIES, PARTICULARLY “AT-RISK” COUNTRIES

- Ensure appropriate investment in local-language resourcing throughout the world, with a particular emphasis on proactively resolving existing inequalities that disproportionately impact non-English speaking, Global Majority countries.
- Ensure equality and consistency between jurisdictions in terms 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

- Prohibit targeted advertising on the basis of invasive tracking practices.
- Ensure that access to essential digital services and infrastructure such as the Facebook platform and other social media platforms are not made conditional on the ubiquitous surveillance of platform users. Access to digital services and infrastructure must be based on the free and informed consent of platform users. This will require enacting and/or enforcing legislation to guarantee people a “right not to be tracked” by advertisers and other third parties. Subscription models which force users to pay in order to avert invasive surveillance undermine freely given consent and therefore should not be permitted.
- Introduce obligations for social media 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 consent for opting in being freely given, specific, informed and unambiguous. The collection and use of inferred data (for example, recommendations based on watch time and likes) to personalize ads and content recommendations should be banned. Rather, users should be in control of which signals or declared interests they want the platform to factor into the shaping of their feed. For those who prefer a feed based on personalized recommendations, they must be given the option to communicate personal interests to the platform based on specific, freely given and informed consent.
- Ensure adequate investment in independent oversight, monitoring and enforcement of regulations governing the technology sector.
- Ensure that truly independent national data protection regulators have adequate resources and expertise to meaningfully investigate and sanction violations by social media companies in line with international human rights law and standards. Governments must also ensure effective individual and collective redress mechanisms.

APPENDIX 1

META'S RESPONSES TO AMNESTY INTERNATIONAL



October 25, 2023

Ms Rasha Abdul Rahim
Programme Director
Amnesty Tech
Amnesty International
1 Easton Street,
London, WC1X 0DW,
United Kingdom
by email

Dear Ms. Abdul-Rahim,

Thank you for reaching out regarding Amnesty International's intention to publish the draft report mentioned in your 11 October 2023 letter (Amnesty reference: TC AFR 25/2023.4613).

In these circumstances, I regret I'm unfortunately not in a position to comment further.

The content of the draft report pertains to matters currently subject to litigation involving Meta Platforms, Inc. ("Meta") before the High Court of Kenya (Petition No. E541 / 2022), in which Amnesty International is involved as an Interested Party. As such, given rules against parties to a dispute publicizing matters that are under consideration by Kenyan courts, Meta respectfully declines to provide detailed comments on the allegations outside the ongoing proceedings and reserves its rights.

We fundamentally disagree with the conclusions Amnesty International has reached in the draft report as they are in fact unproven allegations made by Petitioners in the above referenced litigation.

The allegations of wrongdoing against Meta in the draft report ignore important context and facts including, for example,

- (i) that Ethiopia has been one of our highest priorities for country-specific interventions, and
- (ii) Meta's significant and ongoing efforts to curb violating content on Facebook related to Ethiopia.

Details of these are available in our Letter to Amnesty International of 18 September 2023, as well as in our 2021 [annual Human Rights Report](https://humanrights.fb.com/wp-content/uploads/2023/09/2021-Meta-Human-Rights-Report.pdf), available online at <https://humanrights.fb.com/wp-content/uploads/2023/09/2021-Meta-Human-Rights-Report.pdf>.

Yours sincerely,



Miranda Sissons
Director, Human Rights Policy

Attached: Letter to Amnesty International of 18 September 2023



September 18, 2023

Ms Rasha Abdul Rahim
Programme Director
Amnesty Tech
Amnesty International
1 Easton Street,
London, WC1X 0DW,
United Kingdom
by email

Dear Ms Abdul Rahim,

Thank you for your letter of July 18.

First, we believe it may be useful to begin by including information we published May 2022 in response to the Oversight Board, publicly available at [Meta Q1 2022 Quarterly Update on the Oversight Board](#):

The global Human Rights team has worked as an integral part of the At Risk Countries (ARC) program since 2019. An overview of our approach to countries at risk can be found in our [Transparency Center](#). The team has embedded standards and principles from the UN Guiding Principles on Business and Human Rights (UNGPs) into relevant prioritization and analytical frameworks, which serve as ongoing due diligence tools. Shortly after our global Human Rights team was created, Ethiopia's risk level was re-evaluated, a third party fact checking assessment program was established, and other colleagues were trained on atrocity crime risks and frameworks. Human Rights staff have also participated in Ethiopia risk management processes on an ongoing basis since 2019.

Since early 2020, due diligence has included field-based information gathering and research, focused stakeholder engagement, and formal due diligence from

the Business and Human Rights Group in partnership with our Africa Public Policy team. An updated Human Rights Impact Assessment (HRIA) focused on risk mitigation ahead of Ethiopia's elections was finalized in June 2021.

In 2022, members of the cross-functional team of Meta staff visited Addis Ababa for three days to meet with a wide array of stakeholders that included: the Prime Minister's Office, various government agencies, human rights activists, CSOs, our Trusted Partners, and independent institutions including the Ethiopia Media Authority and the Ethiopia Human Rights Commission. The team has also engaged with new stakeholders about our Community Standards, digital security, safety and privacy standards, connectivity work, and our socio-economic impact initiatives.

This extensive and multi-layered due diligence and stakeholder engagement has informed our approach to content moderation and product development for the region. Our human rights teams continue to support Content Policy teams to understand the humanitarian situation, codewords, and slurs being used on the ground. Relevant members of our Content Policy teams have also received law of armed conflict training. We can now review content in the top four languages spoken and those central to the conflict (Amharic, Oromo, Somali, Tigrinya). Additionally, we have created solutions that make it easier for Ethiopians, as well as specialized international and local human rights and civil society organizations, to notify us when they see potentially violating content, so we can investigate it for possible violations. As stated previously in response to the Board's recommendation, we will continue to conduct thorough human rights due diligence and dynamic risk management processes in Ethiopia, and share insights and actions from our due diligence that aligns with the board's recommendations, our Human Rights Policy, and the UNGPs.

We note we concluded another trip, similar to that described above, in September 2023.

Second, while privacy and safety risks precluded the full publication of our Ethiopia due diligence, we released our first ever annual [human rights disclosure report](#) in mid 2022. The report included our approach to due diligence, insights and actions from our ongoing efforts to identify, prevent, and mitigate human rights risks around the world; UNGP-informed factors to determine countries at risk and a brief overview of our interventions in Ethiopia as outlined below.

You may find pages 65-69 highly relevant. We're excerpting the majority of it here for your convenience:

Integrated Analytical Approaches Countries at Risk

In identified countries (and priority elections), we take a more customized comprehensive due diligence, risk mitigation and response approach — acting quickly to initiate the process to remove content that violates our policies and taking protective measures, including deploying country-specific support. Since 2018, this work has been led by teams with expertise in human rights, hate speech, women and child protection, misinformation, and polarization. Many have lived or worked in high-risk countries and speak relevant languages.

Beginning in 2019, we developed a human rights due diligence process for reviewing and prioritizing countries with the highest risk of offline harm and violence, every six months. To assess long-term conditions on the ground, a cross-functional team of data scientists, political scientists and regional experts review quantitative and qualitative data twice a year drawn from over 60 sources including Varieties of Democracy (V-Dem), Uppsala Conflict Data Program, the United States Holocaust Memorial Museum's Early Warning Project, the Armed Conflict Location & Event Data Project, and the World Bank. This data can include information regarding civic participation and human rights, societal tensions and violence, and the quality of relevant information ecosystems.

UNGP-Informed Factors to Determining Countries at Risk:

Long-term conditions and historical context: We rely on regional experts, platform data and data from more than 60 sources to assess the long-term conditions on the ground. • How much the use of our products could potentially impact a context: We prioritize countries based on a number of factors, including: where our apps have become most central to society, such as in countries where a larger share of people use our products; where there is potential for an increase in offline harms; and where social media adoption has grown. • Current events on the ground: We give special consideration to discrete events that might magnify current societal problems, such as local risk or occurrence of atrocity crimes, polarizing elections, episodes of violence, and COVID-19 vaccination and transmission rates

Guided by the above, we provide specialized support and focus teams on countries most at risk by working to drive human rights to the front of our product, policy, operations, and corporate decision-making. Our focus is

threefold: remove content that violates our policies, respect people's right to free expression, and help to keep people safe both online and offline.

As part of our work, we took the following actions: • Understand and engage with local contexts and communities to ensure our enforcement accounts for local context. We expanded our global network of third-party fact-checkers, invested significant resources in more than 30 countries with active conflict or societal unrest, and together with UN partners and dozens of local and global NGOs, we developed programming to make online engagement safer. • Develop and revise policies to prohibit harmful content and behavior by constantly evaluating and refining our policies to address evolving nuances of hate speech, identify groups at heightened risk of violence and human rights abusers, or the potential for rumors and misinformation to contribute to physical harm, particularly in countries where ethnic and religious tensions are present. • Rapidly enhance our technology and enforcement to help keep our community safe when the risk of impact is greater, i.e. in advance of elections and during periods of heightened unrest. For example, we may reduce or restrict distribution of content from accounts previously found to post content violating our policies, while our teams investigate. Also, we may offer additional features, e.g. providing people in Afghanistan following the Taliban takeover, the ability to lock their profile as an extra layer of privacy, security and protection. • Improve our moderation across languages by adding more expertise.

Case Study

b. Ethiopia

As of this report's publication, Ethiopia is undergoing a conflict that began in November 2020 between the government of Ethiopia and forces in the northern Tigray region. As the turmoil spreads, the situation remains volatile. Ethiopia is an especially challenging environment to address these issues, in part because there are multiple languages spoken in the country. Meta and its platforms are key avenues for communication in Ethiopia. Diligence: Since early 2020, human rights due diligence for Ethiopia has included field-based information gathering and research; focused stakeholder engagement; and formal due diligence. In 2021, Meta commissioned forward-looking human rights due diligence to help identify existing and prospective human rights risks and inform internal decision-making. We engaged activists, human rights defenders, journalists, civil society groups throughout Ethiopia and the diaspora, adding to the insights gained regularly from human rights groups, NGOs, local civil society organizations, and other regional experts.

Action: Using learnings from our due diligence, Meta increased efforts to tackle harmful content (e.g., violence and incitement, hate speech, coordinated inauthentic behavior, misinformation and harm), and addressed elections-related risks and content involving dangerous organizations and individuals. We also: • Improved and simplified our reporting tools to make it easier for Ethiopian users to report possible violating content, so we can investigate. • Rolled out a media literacy campaign, aimed at educating and informing people on how to detect potential false news, and ran billboard advertising campaigns across Addis Ababa, the first of its kind in Africa, focused on informing and educating people on how to stay safe online and use social media responsibly.

Partnered with the Center for African Leadership Studies to implement "My Digital World," a series of live webinars on topics such as online safety, privacy, digital citizenship, news and media literacy. • Established dedicated reporting channels for specialized international and local human rights and civil society organizations to quickly review problematic content they identify for possible violations. • Launched a Lock Your Profile feature (as in Myanmar) for Facebook, providing additional safety, privacy and security for our on-the-ground users. • Increased our investment in proactive detection technology that helps identify violating content such as hate speech before it is reported, including adding in Amharic and Oromo capabilities. • Hired more Ethiopian nationals with native language skills to monitor and respond to threats in real time, and onboarded more Trusted Partners to provide contextualized advice.

Third, we've also publicly shared other updates that address many of your questions. Relevant links include:

[Removing-coordinated-inauthentic-behavior-from-Ethiopia](#)

[Our Approach to Maintaining a Safe Online Environment in Countries at Risk | Meta](#)

[How Facebook is Preparing for Ethiopia's 2021 General Election](#)

[An Update on Our Longstanding Work to Protect People in Ethiopia | Meta](#)

Meta Response to Oversight Board on the Raya Koo case:
[Oversight Board Selects a Case Regarding a Post Discussing the Situation in Ethiopia](#)

Updated Response to Oversight Board on the Raya Koo case:

[Meta Q1 2022 Quarterly Update on the Oversight Board](#)

[Meta Human Rights Report 2022](#) (See pages 65-69)

[Introducing 22 system cards that explain how AI powers experiences on Facebook and Instagram](#)

We do hope you find these useful. We do regret that current litigation makes it impossible to discuss related issues further.

Yours sincerely,

A handwritten signature in black ink, appearing to read 'Miranda Sissons'.

Miranda Sissons

Director, Human Rights Policy

APPENDIX 2

SAMA'S RESPONSE TO AMNESTY INTERNATIONAL

Disclosure:

Sama disputes the claims made against the company in court and is committed to finding a resolution that satisfies all parties.

Sama made the strategic decision to exit all content moderation work to focus on our core competency of computer vision data annotation solutions. That exit was complete as of March 2023.

Content moderators who worked for Sama in Kenya complained of constant pressure from managers to work at speed, with a requirement to deal with each ticket or "item" in 50 to 55 seconds.

There was no required quota. Goals were focused on quality, not quantity. Content moderators did not have targets on the number of tasks per day, and there were no constraints on handling time for the specific tasks.

Given the nature of the work, we allowed moderators to take breaks as and when they needed it. Sama capped content moderators' work weeks at 37.4 hours, requiring content moderators to take 1.5 hours of wellness and meal breaks per day, and offering 1:1 on-demand counseling sessions 24/7.

Many content moderators previously employed by Sama reported suffering from mental health issues such as depression, anxiety, post-traumatic stress disorder and suicidal ideation after spending hours a day reviewing content which contained graphic violence. A former content moderator who spoke to Amnesty International said that the job had seriously affected their psychological wellbeing, and that they had witnessed colleagues struggle to trust people due to the disturbing nature of the content they had to engage with in their work.

We understand that the nature of this work is challenging. That is why all content moderators had full access to both group and individual wellness counseling via Sama's professionally trained and licensed mental health therapists, all of whom are fully credentialed and accredited through the Kenya Counseling and Psychological Association (KCPA).

As is the policy across Sama, anyone can access wellness at any time, either with the counselors we provide onsite or outside counselors through the company's medical and wrap-around benefits including a 24X7 hotline with licensed professionals. All individuals participated in mandatory sessions and a large percentage accessed additional services on demand. The medical benefits include reimbursement for professionals such as psychiatrists if required.

Many of the content moderators employed by Sama have reported that they did not fully understand what they were signing up for when they took the job, with some believing that they would be doing the equivalent of a customer service role. A former content moderator for a social media platform who was employed by Sama told Amnesty International that when he realised what the job truly entailed, he felt 'devastation', which was made worse by the 'traumatic conditions' of the working environment.

We clearly disclose the nature of content moderation work in a phased approach during the interview process. After initial vetting with a counselor, candidates are introduced to the nature of content moderation work and nature of the job, including the identity of the client. If candidates proceed, they are introduced to sample content. After viewing the sample content we request candidates to consent to proceed with the hiring process.

In addition to clearly disclosing the work to be done and sharing sample content during the interview process, we ensured that the candidates passed our resiliency evaluation before joining the company. Post hire, we consistently looked for any behavior changes and were ready to intervene if the work seemed to have a persistent adverse effect. Mandatory individual psychological well-being sessions were scheduled for each content moderator in Sama, facilitated by trained counselors. We also held group sessions and general mental wellness programs for ongoing well-being management. A trained counselor was available 24/7 for on-demand individual/ group wellbeing support. Post Sama, staff have had access to Sama-paid psychological support for 12 months.

**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



www.facebook.com/AmnestyGlobal



[@Amnesty](https://twitter.com/Amnesty)

“A DEATH SENTENCE FOR MY FATHER”

META’S CONTRIBUTION TO HUMAN RIGHTS ABUSES IN NORTHERN ETHIOPIA

In November 2020, a brutal conflict broke out in northern Ethiopia. In 2021, whistle-blower Frances Haugen said that the Facebook platform was “literally fanning ethnic violence” in places like Ethiopia.

This report is based on an investigation into Meta’s role in the serious human rights abuses perpetrated against the Tigrayan community between 2020 and 2022. It reveals the devastating impacts that the Facebook platform’s surveillance-based business model had in the context of Ethiopia’s armed conflict. It shows that, despite the company’s claim to have learned the lessons of its contribution to the atrocities against the Rohingya in 2017, many of the same systemic failures occurred again in Ethiopia.

Despite a recommendation from its own Facebook Oversight Board to conduct a human rights impact assessment in Ethiopia, Meta has to date failed to adequately engage with its contribution to serious human rights harms in the Ethiopian context. However, Amnesty International’s analysis highlights the urgency of this issue, and underlines the necessity of wide-ranging reforms and meaningful regulation to ensure that Meta does not continue to contribute to human rights abuses in Ethiopia, or in any other conflict-affected setting.