

ESSAY

Responsibility and Sustainability of Social Media Content Moderation

Dev Roychowdhury^{1,2*} and Kibrom Berhane Gessesse³

¹Faculty of Education, The University of British Columbia, Vancouver, Canada

²DR ACADEMY, Melbourne, Australia

³Interdisciplinary Studies Graduate Program, The University of British Columbia, Vancouver, Canada

*Corresponding Author: info@drdevroy.com

The advent of the internet and subsequent rise of social media platforms in the past two decades has led to the inevitable discourse around the technical, environmental, cultural, economic, and social responsibility and sustainability of these platforms and the role they play in an increasingly complex world. Although the merits of social media platforms in making the world more connected are indubitable, the far-reaching detrimental impact and consequences of these social media platforms on local, national, and regional sustainability and stability have received scant attention. In this article, we consider how one of the social media platforms, Facebook, has been complicit in playing a pivotal role in fuelling sectarian violence and hatred

in the country of Ethiopia. By focusing on the Ethiopian scenario, we aim to address the repercussions of negligent use of social media platforms, especially in developing countries in the Global South. More specifically, we call for an urgent reflection on social media content moderation policies and practices. We recommend relevant and practical suggestions and further argue that the proliferation of social media platforms must be balanced with social and corporate responsibility and sustainability.

Keywords: social media; content moderation; Facebook; sustainability; sectarian violence; ethnic violence; hatred; discrimination

Social media companies have been criticised for being negligent in moderating content in the Global South (Stremlau, 2022). One of the biggest social media platforms, Facebook, has particularly been blamed for inciting hatred, ethnic violence, and killing, especially in developing countries such as Ethiopia (Jackson et al., 2022; Mackintosh, 2021; Scheck et al., 2021). As a result, Facebook's parent company, Meta, has been sued for US \$1.6 billion for intensifying ethnic violence in Ethiopia

(Amnesty International, 2022; Perrigo, 2022). CNN revealed that despite ranking Ethiopia in its highest priority tier for countries at risk of conflict, coupled with the repeated alarms sounded by some of its employees, Facebook did little to curtail the deluge of inflammatory content on its platform, which is especially disturbing considering that Facebook was well aware of the fact that militia groups in Ethiopia were using the platform to incite violence against ethnic minorities (Mackintosh, 2021). A former Facebook employee-turned-whistle-blower Frances Haugen, for instance, testified in US Congress that Facebook should be held accountable for spreading content that instigated violence and failing to halt fake news in Ethiopia and Myanmar (Akinwotu, 2021). In a nutshell, Facebook's "lack of investment in Ethiopia allowed for problematic content to be widely disseminated" (Crisis Group, 2023). Facebook has been accused of failing to build necessary systems to detect misinformation or hate speech, with only one fact-checker in Ethiopia to review social media content in Amharic and English languages (Mackintosh, 2021).

It is important to note that Ethiopia represents religious, linguistic, and ethnic diversity, with at least 4 major religious groups, 15 languages, and 10 ethnic groups, among others (CIA World Factbook, 2023). Over the past several years, Ethiopia has endured lengthy conflict-laden periods (Skjerdal & Gebru, 2020). Particularly since 2016, it has been ravaged by political extremism and perpetual ethnic and armed conflicts. The recurring internal ethnic and political conflicts coupled with a lack of public trust in the mainstream media made social media platforms, such as Facebook, among the leading information sources and fertile grounds to seed polarized ethnic and political discourses (Kene & Feyissa, 2020). The complexity of Ethiopia's ethno-political landscape, as such, presents a unique context to critically investigate how Facebook is leveraged for sectarian rhetoric and political extremism in an ethnically divided society occupied with protracted internal conflicts. The increasing popularity of Facebook and other social media platforms in the country allowed legions of social media activists and personalities to influence the ethno-political discourse, which often incites further ethnic divisions and worsens violence (Internews, 2023; Kene & Feyissa, 2020).

"Social media content is driven by the social [and political] context in which it is created" (Erbschloe, 2017, p. 13). The social, multi-cultural, and multi-ethnic milieu in Ethiopia, among others, is characterised by ethnic and political hostilities, uncritical

loyalty to ethnicized political parties, a dearth of informed deliberation, and a large percentage of the youth population with a high demand for access to information (Chekol et al., 2021). As such, despite their potentially positive role, social media platforms can be morphed into tools to stir violence. It is argued that social media platforms may be leveraged as the best weapon to influence narratives, especially during a crisis (Matejic, 2015; Singer & Brooking, 2018). During Ethiopia's two-year Tigray War (2020-2022), each party that took part in the war, including the central government of Ethiopia, Tigrayan forces, regional militia groups, and the neighbouring country of Eritrea have been accused of using social media platforms, such as Facebook, to push narratives that resulted in deaths of hundreds of thousands, displacement of millions, and property destruction worth billions of dollars (Philling & Schipani, 2023; Pohjonen, 2022; Stremlau, 2022).

As Stray and colleagues argue, social media platforms often “respond to conflict through content moderation” (Stray et al., 2023). Algorithms may be designed and deployed to amplify certain types of messages. Recent platform experiments, reports of content creators, and experiences of people who have been in conflicts show that social media platforms such as Facebook usually incentivise divisive and potentially violence-laden speeches (Proctor, 2021). And, therefore, the responsibility of big technology companies in moderating online content in countries such as Ethiopia have been underscored by many concerned researchers and activists (Roychowdhury, 2023; Stremlau, 2022). Considering the 2021 Facebook leaks along with whistle-blower Frances Haugen's testimony, it is evident that the company has been aware of the harmful impacts its platform causes. Haugen described the role of Facebook in Ethiopia's Tigray war, in particular, as ‘terrifying’ and ‘literally fanning ethnic violence’ (Akinwotu, 2021; Hao, 2021). Haugen further indicated that she did what she “thought was necessary to save the lives of people, especially in the global south, who [she] think[s] are being endangered by Facebook's prioritization of profits over people” (Milmo, 2021). As the Facebook Papers obtained by *The Washington Post* further reveal, Facebook “privately and meticulously tracked real-world harms exacerbated by its platforms, ignored warnings from its employees about the risks of their design decisions, and exposed vulnerable communities around the world to a cocktail of dangerous content” (Lima-Strong, 2021). Mackintosh (2021) reported that “Facebook employees repeatedly sounded the alarm on the company's

failure to curb the spread of posts inciting violence in 'at risk' countries like Ethiopia, where a civil war has raged for the past year,” but the company did little to stop it. The leaked documents and the whistle-blower’s testimony affirm the discrimination and chronic problems of content moderation in peripheral countries (such as Ethiopia). According to Stremlau (2022), although 90% of Facebook users are outside North America, a very insignificant percentage of moderators focus on the Global South. In the case of Ethiopia, for instance, less than 5% of identified hate speech content is proactively removed, a far cry from Facebook’s claim of 90%. As the documents submitted by Frances Haugen to the United States Securities and Exchange Commission (SEC) indicate, it is only possible to remove approximately 3-5% of hate speech and less than one percent (or only 0.6%, to be precise) of violent and inciting content. A key factor contributing to this difficulty is the substantial challenge associated with incorporating semantic meaning categories into the decision-making process at the content level, “especially when the process needs to be localized across languages and have rudimentary interpretation of intent” (see The Facebook Papers, 2021b). This shows that content moderation in countries like Ethiopia needs a lot of work in terms of technological support, willingness, contextual understanding, perception, and dynamic partnership between people and algorithms. A recent study showed that 96% of Ethiopian respondents reported that they experienced disinformation and misinformation on Facebook and other online platforms (Internews, 2023). Content moderation in the Global South, therefore, seems to be a willingness and capacity issue of the giant technology companies. That is, Facebook has the financial resources and capacity to hire and train content moderators and other technicians to enhance the quality of its services in conflict-prone countries, such as Ethiopia.

Despite Facebook’s claim of implementing an unprecedented third-party fact-checking programme (Pelley, 2021), in the case of Ethiopia, the company has language competency only for two languages out of more than fifty languages spoken (Stremlau, 2022). This means Facebook’s content moderation only applies to a very few dominant couple of languages in the country, whilst discounting a large proportion of other languages. This is a typical example of content moderation in the Global South, particularly in Africa, where such policies and practices are extremely limited and

ineffective. The big technology companies focus on the performance of their platforms in the developed nations whilst the same platforms are complicit in fanning conflicts in the developing countries (Mackintosh, 2021; Njanja, 2023; Roychowdhury, 2023; Scott, 2021). In the case of Facebook, this argument is in line with the allegations detailed in the leaked documents concerning user engagement practices and content moderation. Despite Facebook teams identifying and developing ways to combat misinformation and violence, the company's decision-making on content policy is routinely influenced by political considerations. This highlights an inherent conflict of interest within the technology giant (see The Facebook Papers, 2021b).

WHY IS SOCIAL MEDIA A SUSTAINABILITY ISSUE?

The social media issue in the Global South (such as countries like Ethiopia) is a sustainability issue because it affects the social, economic, and environmental dimensions of the regional development. Socially, the proliferation of hate speech, misinformation, and incitement of violence on Facebook and other social media platforms undermines the social cohesion, peace, and human rights of the diverse population. It also erodes the trust and confidence in the democratic institutions and processes that are essential for a stable and inclusive society. Economically, social media-fueled conflicts disrupt the livelihoods, markets, and infrastructure of countries like Ethiopia, which is one of the fastest-growing economies in Africa. It also hampers the potential for regional integration and cooperation, as well as the attraction of foreign investment and tourism. Environmentally, the social media-induced violence exacerbates the existing challenges of climate change, food insecurity, and displacement that affect millions of people. It also poses a threat to the natural resources and biodiversity of the countries, which are vital to their resilience and adaptation.

Therefore, addressing the social media issue in Ethiopia is not only a matter of corporate responsibility and accountability, but also a matter of sustainability and development. It requires a holistic and collaborative approach that involves the social media companies, the government, the civil society, the media, and the users themselves. It also requires a recognition and respect for the cultural, linguistic, and ethnic diversity of the country, as well as the historical and political contexts that shape its realities. By

doing so, the social media platforms can become a positive force for social change, rather than a catalyst for social chaos.

DISCUSSION

In summary, we have highlighted the role of Facebook in the Global South by focusing on the Ethiopian case. As the discussion reveals, despite the merits, Facebook as a technology, communication, and social media platform has the potential to fan hatred, discrimination, and violent incidents. Hence, in order to make social media platforms a more viable, sustainable, and positive environment for users, we argue that there needs to be a strategic and systematic shift in how social media companies tackle content moderation. Therefore, we suggest the following solutions, which focus on literacy, technicality, and context-sensitive practice.

Promote Digital Literacy to Empower Citizens

Given the cultural diversity, population size, and the high number of daily social media users in the world, with special consideration for high-risk and conflict-affected areas in the Global South, it will be challenging to execute effective content moderation with limited human resources. Hence, having open access online social media literacy courses and trainings for users will be a viable solution. This will help users to assess and appraise social media content by themselves critically. Social media literate users can effectively spot, block, and report trolls (users who intentionally inflame incidents), and other malicious users.

To minimize harm, the so-called 'frictionless design' employed by Facebook and other social media platforms, which relies on users making subconscious and automatic decisions (Narayanan, 2023), requires users to exercise judgment and the ability to resist their impulses. This demands a high level of media literacy. Social media literacy, in particular, will be vital for users to understand how platform algorithms work, empowering them in their social media interactions and demystifying the mythologies of algorithms with which they interact for several hours a day on average (Statista, 2024). In the case of Ethiopia, various prominent stakeholders can assume responsibility for enhancing media literacy, including non-governmental organizations focused on youth education and social media affairs, public universities, and the Youth Affairs Office under

the Federal Democratic Republic of Ethiopia's Ministry of Women and Social Affairs. For instance, the Centre for the Advancement of Rights and Democracy (CARD), a non-profit organization based in Addis Ababa, Ethiopia, aims to enhance media literacy among practitioners and users through capacity building to counter disinformation, thereby fostering a vibrant and safe media environment (CARD, 2023). Other key stakeholders should adopt and implement similar practices.

Improve Technical Systems to Create Robust But Responsible Social Media Platforms

Social media companies have the responsibility and capacity to code and design algorithms and software programs that would help fact-check information and authenticate users to create a more real and viable social media ecosystem. For instance, enhanced algorithms may target, identify, and dismantle malicious spambots, trolls, and fake accounts to ensure that platforms are free from such agents that may create and spread false, misleading, or harmful content.

Technology professionals and writers argue that "just as algorithms can be trained to predict who would click what ad, they can also be trained to predict who would like or share what post" (Hao, 2021). Therefore, Facebook, for instance, can control engagement-based content ranking by integrating it with effective integrity and security systems instead of focusing solely on machine-learning models or algorithms that maximise engagement by favouring controversy, misinformation, and extremism. Due to algorithmically supercharged content recommender systems, it is impossible for human reviewers alone to keep up with the volumes of content (Narayanan, 2023; Yibeltal & Muia, 2023). To address this, Facebook should ensure seamless, effective, and dynamic human-algorithmic cooperation for content moderation. Moreover, Facebook should enhance its integrity system (a system designed to address incendiary content on the platform) and content policy in many languages other than English. In the case of Ethiopia, for instance, Facebook's integrity system supports only two of the six widely used social media languages. Therefore, Facebook should improve its ability to enforce its content policy in multiple languages. Meta, Facebook's parent company, has indeed conducted an internal audit of its algorithms' operations. However, the extent of changes is still unclear (Crisis Group, 2023), as the company could not publish its findings for security and confidentiality reasons (France 24, 2024).

Employ Context-Sensitive Mechanisms to Moderate Social Media Content

Content moderation is a contextual and reactive process that may become arduous to manage and/or moderate. For instance, it may be difficult to discern and police certain unambiguous or culturally specific forms of harmful content from the more obvious forms of hate speech or violence, which is a massive undertaking. Furthermore, in nations like Ethiopia, the intricate ethno-political environment and rapidly evolving conflict dynamics require social media companies to prioritize and invest significantly in content moderation efforts (Crisis Group, 2023). Hence, proactive and context-sensitive mechanisms that can prevent the creation and sharing of harmful content would be pivotal. That is, parallel to algorithmic selection, human adjudication must be able to successfully differentiate between useful and positive viral content, and scrupulous or toxic misinformation that attracts virality and high engagement due to their divisiveness and maleficence. Protective mechanisms to moderate social media content must include the hiring of language and cultural experts that would comprehensively cover the richness and diversity existent in the Ethiopian milieu. We argue that Ethiopia's experience with social media is distinct and intricate compared to many other nations, particularly those with well-established democracies. This distinction arises because the interactive capabilities of Facebook and other social media platforms preceded Ethiopia's engagement with robust and rational political discourse as we know it today. The complexity stems from Ethiopia's intricate ethno-political history, which has transitioned through various ideologically divergent political systems, including imperial rule (1930-1974), Marxist-Leninist regime (1975-1991), ethno-federalism (1991-present), and the more recent tumultuous period stemming from the rapid political reform (since 2018). These transitions have been accompanied by political repression, human rights violations, and economic instability. Consequently, if not carefully managed, as witnessed during the Tigray War (2020-2022) and the subsequent ongoing conflict in the Amhara region between the central government and the Amhara militia known as Fano, any online practices involving misinformation or disinformation can easily ignite violent conflicts.

It is critical to note that we are not arguing against the use of social media platform or suggesting that they, by their very nature, are drivers of violence. However, situated in a broader context of violence and volatile political context, social media platforms can

inflame, catalyse, and exacerbate conflicts. Therefore, as a platform of social interaction and connectivity, Facebook has a considerable stake in influencing social sustainability. Mutually beneficial engagement between the company and its users is, therefore, imperative. Our comments on the negative impacts of Facebook are not here to deny the plethora of positive functions of Facebook or any other social media platform. It is rather to solidify the argument that Facebook (and other social media platforms alike) should strive to create and connect healthy and livable communities to enhance social sustainability.

We would also like to point out that our discussion is not just limited to Ethiopia. The experiences and incidents we discussed here affect most developing countries (and even developed countries). For instance, Facebook has been found to be complicit in a number of cases, such as the Myanmar coup, 2020 US Presidential Election, Middle Eastern countries (e.g., Afghanistan, Iraq, Syria, Yemen), to name a few, where it has been found that the platform has repeatedly failed to protect its users, tackle hate speech targeting minority groups, and employ enough local staff to extirpate religious sectarianism (Mackintosh, 2021; Njanja, 2023; Roychowdhury, 2023; Scott, 2021). For instance, during the US Capitol Hill violence on January 6, 2020, reports on Instagram and Facebook peaked at seven times the hourly rate compared to the previous week. Many of the most reported content featured Donald Trump (or videos of him) and advocated for violence and the overthrow of the government (see Cameron et al., 2023). Furthermore, Haugen states that “Facebook’s records confirm that Facebook knowingly chose to permit political misinformation and violent content/groups and failed to adopt or continue measures to combat these issues, including as related to the 2020 U.S. election and the January 6th insurrection, in order to promote virality and growth on its platforms” (see The Facebook Papers, 2021b). Therefore, the adverse impacts of social media usage are not exclusively limited to developing nations; they also significantly influence the sociopolitical landscape in developed countries. In this paper, we considered the Ethiopian case because given the recurring armed conflicts in recent times coupled with the sociopolitical and cultural diversity and complexity, it sufficiently exemplifies what can go wrong in most developing nations if Facebook and other social media platforms are not used or managed properly.

CONCLUSION

Historical tensions and antagonistic political ideals underlay the Ethiopian state (Pohjonen, 2022). We believe that it is within these political and historical contexts that the roles of Facebook in ethnopolitical polarisation should be understood. As the role of Facebook in Ethiopia and elsewhere in other developing countries show, the promises behind social media platforms are not realised, at least, as intended. Such experiences obscure the roles big technology companies can play in the Global South. Owing to language barriers or the limitations of Facebook's ideological stance and algorithms for monitoring hate speech and incitement of communal violence conveyed in Amharic (the national language of Ethiopia) and other local languages, a distorted online discourse will continue to be a threat to Ethiopian national security and political stability. This has far-reaching consequences, not just for Ethiopia, but also for a large proportion of developing countries in the world.

We believe that technology cannot be divorced from the social conditions of its creation and application. So far, the applications of advanced technologies in the Global South appear to assume that technologies can be free from biases of their social creation. This, however, has not been the case as is evident from experiences and incidents in many parts of the Global South. Hence, social media's overall functioning and content moderation strategies should be designed according to the contexts where they are deployed. The foundational epistemic position that these social media companies seem to be founded or appear to solely focus on boosting algorithmic virality where content dissemination systems prioritize surveillance-based profit-only business models and engagement-centric operations that normalize hate, violence, and discrimination, at the cost of constructive discussions, positive interactions, and even human lives.

Even so, we do note that it perhaps would not be completely fair to place the sole burden of violence on Facebook or other companies. Internal shortcomings of countries in the Global South, such as digital illiteracy, volatile political systems, complex socio-historical contexts, and weak (media) regulatory environment play significant roles. Despite these contextual factors, adapting the content moderation models of Facebook and other social media platforms according to local needs would benefit both the technology companies and local communities. Without such pragmatic undertakings, mere digital

solutionism or technological fixes to complex socio-political problems of nations in the Global South, such as Ethiopia, will be a myth. In other words, in fragile nations such as Ethiopia, Facebook and other social media platforms possess the potential to influence public opinion and conflict dynamics either positively or negatively. While acknowledging Facebook's role as a space for free speech, resistance against oppression, and civic engagement, we argue that these functions can become distorted within volatile sociopolitical contexts, as demonstrated throughout this paper, ultimately exacerbating existing tensions.

And so, in realising the myriad benefits that social media platforms bring to the increasingly dynamic and complex global world order, we acknowledge that the proliferation of such platforms is inevitable, and indeed a necessity. Nevertheless, considering that social media platforms are intricate systems susceptible to emergent behaviours and dynamic feedback patterns, which consequently result in unpredictable virality and nonlinear behavioural trends, it becomes crucial for these platforms to diligently work towards achieving a delicate equilibrium between their long-term business objectives, stakeholder concerns, and the broader societal responsibility and sustainability factors. It is more crucial than ever for these platforms to enforce meaningful systems to genuinely serve the communities they profit from.

References

- Akinwotu, E. (2021). Facebook's role in Myanmar and Ethiopia under new scrutiny. *The Guardian*. Accessed from <https://www.theguardian.com/technology/2021/oct/07/facebooks-role-in-myanmar-and-ethiopia-under-new-scrutiny>
- Amnesty International. (2022). Meta sued for 1.6 billion USD for fueling Ethiopia ethnic violence. Accessed from <https://www.amnesty.org/en/latest/news/2022/12/kenya-meta-sued-for-1-6-billion-usd-for-fueling-ethiopia-ethnic-violence/>
- Cameron, D., Wodinsky, S., DeGeurin, M., & Germain, T. (2023). *Facebook Papers Directory*, Gizmodo. Accessed from <https://gizmodo.com/facebook-papers-how-to-read-1848702919>
- Centre for the Advancement of Rights and Democracy (2023). *The Beginning of the Long Journey: In Pursuit of Promoting Ethiopia's Democratization*. <https://www.cardeth.org/sites/default/files/THE%20BEGINNING%20OF%20THE%20LONG%20JOURNEY.pdf>.
- CIA World Factbook. (2023). Ethiopia. Accessed from <https://www.cia.gov/the-world-factbook/countries/ethiopia/>

- Chekol, M. A., Moges, M. A., & Nigatu, B. A. (2021). Social Media Hate Speech in the Walk of Ethiopian Political Reform: Analysis of Hate Speech Prevalence, Severity, and Natures. *Information, Communication & Society*.
<https://doi.org/10.1080/1369118X.2021.1942955>
- Crisis Group (2023). *What Facebook Does (and Doesn't) Have to Do with Ethiopia's Ethnic Violence*. Accessed from <https://www.crisisgroup.org/africa/horn-africa/ethiopia/what-facebook-does-and-doesnt-have-to-do-ethiopia-ethnic-violence>
- Erbschloe, M. (2017). *Social Media Warfare: Equal Weapons for All*. CRC Press, Taylor & Francis Group.
- France 24 (2024). *Facebook accused of inciting ethnic violence in Ethiopia*. Accessed from <https://www.france24.com/en/tv-shows/focus/20240131-facebook-accused-of-inciting-ethnic-violence-in-ethiopia>
- Hao, K. (2021). The Facebook whistleblower says its algorithms are dangerous. Here's why. *MIT Technology Review*. Accessed from <https://www.technologyreview.com/2021/10/05/1036519/facebook-whistleblower-frances-haugen-algorithms/>
- Internews. (2023). Ethiopia digital media information ecosystem assessment. *Internews*. Accessed from <https://internews.org/wp-content/uploads/2023/08/Ethiopian-Digital-Media-IEA-edited-Final-SinglePage.pdf>
- Jackson, J., Kassa, L., & Townsend, M. (2022). Facebook 'lets vigilantes in Ethiopia incite ethnic killing'. *The Guardian*. Accessed from <https://www.theguardian.com/technology/2022/feb/20/facebook-lets-vigilantes-in-ethiopia-incite-ethnic-killing>
- Kene, S. M., & Feyissa, S. (2020). Ethiopia's War of Narratives. *Africa is a Country*. Accessed from <https://africasacountry.com/2020/11/ethiopia-war-of-narratives>.
- Lima-Strong, C. (2021). A whistleblower's power: Key takeaways from the Facebook Papers. *The Washington Post*. Accessed from <https://www.washingtonpost.com/technology/2021/10/25/what-are-the-facebook-papers/>
- Mackintosh, E. (2021). Facebook knew it was being used to incite violence in Ethiopia. It did little to stop the spread, documents show. *CNN*. Accessed from <https://www.cnn.com/2021/10/25/business/ethiopia-violence-facebook-papers-cmd-intl/index.html>
- Matejic, N. (2015). *Social Media Rules of Engagement: Why your online narrative is the best weapon during a crisis*. John Wiley & Sons Australia.
- Milmo, D. (2021). Frances Haugen: 'I never wanted to be a whistleblower. But lives were in danger'. *The Guardian*. Accessed from <https://www.theguardian.com/technology/2021/oct/24/frances-haugen-i-never-wanted-to-be-a-whistleblower-but-lives-were-in-danger>.
- Narayanan, A. (2023). *Understanding Social Media Recommendation Algorithms: Towards a better-informed debate on the effects of social media*. Knight First Amendment Institute at Columbia University. Accessed from <https://knightcolumbia.org/content/understanding-social-media-recommendation-algorithms>

- Njanja, A. (2023). Meta pressed to compensate war victims amid claims Facebook inflamed Tigray conflict. *TechCrunch*. Accessed from <https://techcrunch.com/2023/11/08/meta-and-war-victims-in-ethiopia/>
- Pelley, S. (2021). Whistleblower: Facebook is misleading the public on progress against hate speech, violence, misinformation. *CNBC News*. <https://www.cbsnews.com/news/facebook-whistleblower-frances-haugen-misinformation-public-60-minutes-2021-10-03/>
- Perrigo, B. (2022). New lawsuit accuses Facebook of contributing to deaths from ethnic violence in Ethiopia. *TIME*. Accessed from <https://time.com/6240993/facebook-meta-ethiopia-lawsuit/>
- Philling, D., & Schipani, A. (2023). War in Tigray may have killed 600,000 people, peace mediator says. *Financial Times*. Accessed from <https://www.ft.com/content/2f385e95-0899-403a-9e3b-ed8c24adf4e7>
- Pohjonen, M. (2022). An epistemic proxy war? Popular communication, epistemic contestations and violent conflict in Ethiopia. *Popular communication*, 20(3), 236–252. <https://doi.org/10.1080/15405702.2022.2074998>
- Proctor, K. (2021). *Social media and conflict: Understanding risks and resilience*. Mercy Corps.
- Roychowdhury, D. (2023). How social media platforms shape our world: Impacts, challenges, opportunities, and solutions. *Dr Dev Roychowdhury*. Accessed from <https://www.drdevroy.com/how-social-media-shapes-our-world/>
- Scheck, J., Purnell, N., & Horwitz, J. (2021). Facebook employees flag drug cartels and human traffickers. The company's response is weak, documents show. *The Wall Street Journal*. Accessed from <https://www.wsj.com/articles/facebook-drug-cartels-human-traffickers-response-is-weak-documents-11631812953>
- Scott, M. (2021). Facebook did little to moderate posts in the world's most violent countries. *Politico*. Accessed from <https://www.politico.com/news/2021/10/25/facebook-moderate-posts-violent-countries-517050>
- Sjerdal, T., & Gebru, S. (2020). Not quite an echo-chamber: ethnic debate on Ethiopian Facebook pages during times of unrest. *Media, Culture & Society*, 42(3), 365–379. <https://doi.org/10.1177/0163443719895197>
- Singer, P. W., & Brooking, E. T. (2018). *LikeWar: The Weaponization of Social Media*. Eamon Dolan/Houghton Mifflin Harcourt.
- Statista. (2024). *Daily time spent on social networking by internet users worldwide from 2012 to 2024*. Accessed from <https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide/>
- Stray, J., Iyer, R., & Larrauri, H. P. (2023). The algorithmic management of polarization and violence on social media. *Knight First Amendment Institute at Columbia University*. <https://knightcolumbia.org/content/the-algorithmic-management-of-polarization-and-violence-on-social-media>
- Stremlau, N. (2022). Online speech and offline violence: Reflections on the current violence in Ethiopia. *Global Media Journal – German Edition*, 12(2). <https://doi.org/10.22032/dbt.55517>
- The Facebook Papers (2021a). *Facebook Papers Catalogue*. Accessed from <https://facebookpapers.com>

The Facebook Papers (2021b). *SEC Documents*. Accessed from <https://facebookpapers.com/sec-documents/>

Yibeltal, K. & Muia, W. (2023). *Facebook's algorithms 'supercharged' hate speech in Ethiopia's Tigray War*. BBC. Accessed from <https://www.bbc.com/news/world-africa-67275219>

Funding and Acknowledgements

The authors declare no funding sources or conflicts of interest.

Online Connections

Dr. Dev Roychowdhury: <https://www.x.com/drdevroy>
<https://www.linkedin.com/in/drdevroychowdhury>
<https://www.drdevroy.com/>