Date: 10 July 2023   Index number: IOR 40/6986/2023

CONTEMPORARY FORMS OF HATE SPEECH ONLINE

SUBMISSION TO THE SPECIAL RAPPORTEUR ON CONTEMPORARY FORMS OF RACISM, RACIAL DISCRIMINATION, XENOPHOBIA AND RELATED INTOLERANCE

Amnesty International welcomes the call[1] by the UN Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia, and related intolerance ahead of your report on contemporary forms of racism on online hate speech which will examine the different forms of hate speech online, and how it amounts to racial discrimination, hatred and violence, as well as the role of private companies in preventing and addressing online hate speech.

The surveillance-based business models of Big Tech companies such as Google and Meta are the root cause of harmful content proliferation on their platforms. This business model is based on algorithms that amplify hateful and harmful content, which can create echo chambers of hateful content that can result in offline violence, as was the case with atrocities committed against the Rohingya in 2017. Platform companies, such as Meta and Google remain unregulated, and they do not take sufficient measures to prevent harms of their products and hateful content that is spread on their platforms.

THE SURVEILLANCE-BASED BUSINESS MODEL OF BIG TECH PLATFORM COMPANIES:

Access to the internet is now vital to enable the enjoyment of human rights. For more than 4 billion people, the internet has become central to how people communicate, learn, participate in the economy, and organise socially and politically. Yet when these billions participate in life online, most of them rely heavily on the services of just two corporations, namely Google and Meta. Google and Meta control the primary channels that people rely on to engage with the internet. They provide services so integral that it is difficult to imagine the internet without them. Big Tech companies such as Google and Meta's platforms come at a systemic cost. The companies' surveillance-based business model forces people to make a Faustian bargain, whereby they are only able to enjoy their human rights online by submitting to a system predicated on human rights abuse. The surveillance- based business model is an assault on the right to privacy on an unprecedented scale, and along with series of knock-on effects that pose a serious risk to a range of other rights, from freedom of expression and opinion, to freedom of thought and the right to non-discrimination.[2] The surveillance based business model can also lead to the proliferation of content inciting violence and advocacy of hatred, as will be illustrated further below.

Amnesty International has found that the business model of specifically Meta and Google relies on corporate surveillance via the mass harvesting and monetisation of people's data- this is the "...raw facts about our lives and behaviours, and when processed and organised increasingly reveals a huge amount about our innermost thoughts, behaviours and identities."[3]

Data collection is not limited to people signed up to the services or platforms of the companies, but includes virtually anyone who uses the internet, due to the pervasiveness of the company's tracking capabilities across the web. Companies then use algorithmic systems to analyse the data, assign the detailed profiles to individuals and groups, and then predict people's interests and behaviour – essentially inferring and creating new information about people. Access to this

---

[1] https://www.ohchr.org/en/calls-for-input/2023/call-input-online-hate-speech
[2] *Surveillance Giants, previously cited p.5.*
[3] *Surveillance Giants, previously cited, p.9.*

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.

information is then sold to advertisers wishing to reach a defined group of people. To be as attractive as possible to potential advertisers, Big Tech companies compete to offer the best predictions about people, which requires them to expand their datasets and refine predictive algorithms. "This incentivizes the companies to seek more data on more people to expand their operations across the internet, into physical space and, ultimately, across the globe."[4]

This business model is based on ubiquitous surveillance, which reaches well beyond information people may input when signing up to a platform or service (for example, email addresses, birth dates, phone numbers) to include details such as location, search history and app use. Information collected is not limited to data, but also includes meta-data, including "...email recipients, location records, and the timestamp on email and photos".[5] This meta-data allows for increasingly complex analytics and has the potential to infer sensitive details about a person, such as their *sexual identity, political views, personality traits or sexual orientation.*

Big Tech's ubiquitous corporate surveillance is an unprecedented violation of the right to privacy, which is fundamentally incompatible with the companies' responsibilities to respect human rights. "This goes beyond an intrusion into every aspect of our lives online, and in fact threatens our right to shape and define who we are as autonomous individuals in society."[6] This violation of rights shapes our thoughts and creates environments in which hateful content and incitement to commit violence are promoted and proliferated.

HUMAN RIGHTS RISKS AS A RESULT OF THE SURVEILLANCE-BASED BUSINESS MODEL

As explained above, Google and Meta's platforms are underpinned by a set of advanced data analytics systems. "Their algorithmic models are designed to serve a user 'relevant' content (relevancy that is inferred by the companies on the basis of collected data)".[7] For example, the algorithms powering Google Search and Facebook's Newsfeed are continuously trained on vast amounts of user data to serve many different purposes, such as ad targeting and delivery, serving search results, recommending new content, and prompting users to create new content and engage with existing content. To do this, the systems "optimize" to best deliver a specific outcome using highly complex and iterative algorithmic processes that draw correlations and inferences from user data.[8] These algorithmic systems have been shown to have knock-on effects that can result in serious negative impacts on human rights, including privacy, freedom of expression and the right to equality and non-discrimination.[9]

The right to privacy is "an essential requirement for the realization of the right to freedom of expression" and therefore Google and Meta's erosion of the "private sphere" has corresponding direct and indirect impacts on the free development and exchange of ideas.[10] The level of intrusion to privacy by the surveillance business model cannot be considered necessary or proportionate under international human rights law, which establishes that the right to privacy must be guaranteed against arbitrary interferences, "whether they emanate from State authorities or from natural or legal persons". There are three inter-related concepts around privacy: freedom from intrusion in our private lives, the right to control information about ourselves, and the right to a space in which we can freely express our identities. The surveillance-based business model violates the right to privacy on each of these points. Freedom of expression is a collective right, enabling people to seek and receive information as a social group and to "voice their collective views". By their very nature, algorithmic systems impact people as a group as well as at an individual level.

The surveillance-based business model has created an architecture that has not only drastically shrunk and restricted the "private sphere", but at the same time isolated people from one another, as each individual engages with their own highly personalised experience of the internet, uniquely tailored to them based on algorithmically-driven inferences and profiling. This leaves the door wide open to abuse by manipulating people at scale, which includes intruding the thoughts and influencing the actions of people towards hateful content and incitement to violence. [11]

---

[4] *Surveillance Giants, previously cited, p. 12.*
[5] *Surveillance Giants, previously cited, p. 16.*
[6] *Surveillance Giants, previously cited, p. 22.*
[7] *Surveillance Giants, previously cited, p. 27.*
[8] *Surveillance Giants, previously cited, p. 27.*
[9] *Surveillance Giants, previously cited, p. 28.*
[10] *Surveillance Giants, previously cited, 2019, p. 31.*
[11] *Surveillance Giants, previously cited, p. 32.*

CONTEMPORARY FORMS OF HATE SPEECH ONLINE
SUBMISSION TO THE SPECIAL RAPPORTEUR ON CONTEMPORARY FORMS OF RACISM, RACIAL   2
DISCRIMINATION, XENOPHOBIA AND RELATED INTOLERANCE

Amnesty International

META'S ALGORITHMS AND HOW THEIR AMPLIFICATION CONTRIBUTE TOWARDS HATEFUL CONTENT AND INCITEMENT TO COMMIT VIOLENCE

Content shaping algorithms can lead to the creation of social media "echo chambers" whereby individuals are only shown information which reinforces their pre-existing views and beliefs.[12] Meta's use of content-shaping algorithms is designed to maximize engagement, particularly with its newsfeed and recommendations algorithms that facilitate the dissemination of hate speech and harmful content.

Meta's surveillance-based business model incentivizes the spread of maximally engaging, inflammatory content, including advocacy of hatred. The engagement-based ranking of content on the Facebook platform allows the algorithms to amplify speech fans hate speech and ethnic violence. In the Facebook papers that were submitted to the US Congress in a redacted form as part of Frances Haugen's testimony it is stated that there is *"evidence from a variety of sources that hate speech, divisive political speech and misinformation of Facebook and the family of apps are affecting societies around the world. We also have compelling evidence that our core product mechanics such as virality, recommendations and optimizing for engagement are a significant part of why these types flourish on the platform"*.[13]

Meta's content shaping algorithms amplifies hateful and harmful speech in the following ways:

- o **Virality**

One of Facebook's most well-known features is algorithmic virality-whereby certain content is algorithmically boosted in such a way that enables it to reach a far wider audience than it could otherwise reach 'organically' (without being boosted). Meta's algorithmic ranking systems have specific and separate predictions for what they think you would engage with but also what they think you may share so that others may engage with. Outrage and misinformation are more likely to be viral. This means that Meta maximizes engagement to stoke emotive responses. A trove of leaked documents by whistleblower, Frances Haugen revealed that "[a]s long as we continue to optimize for overall engagement and not solely what we believe individual users will value, we have an obligation to consider what the effect of optimizing for business outcomes has on the societies we engage in".[14]

- o **Newsfeed and ranking**

Meta's News Feed algorithm is central to what content is seen by whom on the Facebook platform.[15] It curates a highly personalized user experience based on invasive profiling enabled by the collection of intimate personal data. "Newsfeed ranking is another way Facebook becomes actively involved in these harmful experiences. Of course, users also play an active role in determining what content they are connected to through feed, by choosing who to friends and follow. Still, when and whether a user sees a piece of content is also partly determined by the ranking scores our algorithm assign, which are ultimately under our control."[16] "Actively ranking content in the News Feed and promoting content on recommendations surfaces makes us [Meta] responsible for any harm caused by exposure to that content." "This means, according to ethicists, Facebook is always at least partially responsible for any harmful experiences on the News Feeds. This doesn't owe any flaw with our News Feed ranking system, it's just inherent to the process of ranking... Crucially, all of these algorithms produce a single score, a score Facebook assigns. Thus, there is no such thing as inaction on feed, we can only choose to take different kinds of actions."[17]

- o **R**ecommendations for user behaviour and actions

Meta deploys algorithmic technologies to actively promote and recommend certain actions and behaviours to Facebook users. Recommendations are algorithmically delivered to users on their feeds. Meta's use of content-shaping algorithms to *actively amplify and promote harmful cont*ent bestows a heightened responsibility on the company for any harms caused by that content.[18]

---

[12] Amnesty International, *"Myanmar: The social atrocity: Meta and the right to remedy for the Rohingya"* (ASA 16/5933/2022), 29 September 2022, Myanmar: The social atrocity: Meta and the right to remedy for the Rohingya - Amnesty International, p. 21. See Matt Schissler, "Echo chamber in Myanmar: Social Media and the ideological justifications for mass violence", 2014, Paper for the Australian National University Department of Political and Social Change Research Colloqium, "Communal Conflict in Myanmar:Characteristics, Causes, Consequences. Echo chambers in Myanmar: Social media and the ideological justifications for mass violence - [PDF Document] (vdocuments.mx)

[13] A whistleblower's power: Key takeaways from the Facebook papers- https://www.washingtonpost.com/technology/2021/10/25/what-are-the-facebook-papers/

[14] The Facebook Papers, "We are Responsible for Viral Content", 11 December 2019, p.8" - https://facebookpapers.com/

[15] *Myanmar: The social atrocity, previously cited,* p. 43.

[16] The Facebook Papers, "Facebook and Responsibility", https://facebookpapers.com/; *Myanmar: The social atrocity, previously cited,* p. 43.

[17] The Facebook Papers," Facebook and Responsibility", https://facebookpapers.com/; *Myanmar: The social atrocity, previously cited,* p. 43.

[18] *Myanmar: The social atrocity, previously cited,* p. 43.

CONTEMPORARY FORMS OF HATE SPEECH ONLINE
SUBMISSION TO THE SPECIAL RAPPORTEUR ON CONTEMPORARY FORMS OF RACISM, RACIAL   3
DISCRIMINATION, XENOPHOBIA AND RELATED INTOLERANCE

Amnesty International

o Groups

Meta internal research report recognized that its content shaping algorithms were responsible for the formation of extremists, divisive and politically divisive groups.[19]

IMPACT OF META'S BUSINESS MODEL AND ALGORITHMIC AMPLIFICATION IN FUELLING HATE SPEECH RESULTING IN OFFLINE VIOLENCE AGAINST THE ROHINGYA COMMUNITY

The atrocities experienced by the Rohingya community in Myanmar as a result of Meta's business model along with the algorithmic amplification of hate speech that is a result of a core features of their product, namely Virality; Newsfeed and ranking; Groups; and Recommendations for user behaviour and actions, is thoroughly documented by Amnesty International in its report, 'Myanmar: The Social Atrocity: Meta and the right to remedy for the Rohingya'.

Posts on Meta's platform, Facebook, inciting discrimination against the Rohingya were prevalent and rampant before and during the 2017 atrocities and helped to normalize their dehumanization as an ethnic group in extremely violent and hateful terms, facilitating the rapid spread of Rohingya hatred across the country. The virality of these posts created a sense of urgency and hysteria in finding a "solution" to the "Bengali problem", which created a perceived threat of the Rohingya and building of support for the mass violence which took place against the community in 2017.[20] The content of these posts included the advocacy of hatred with the incitement to commit violence against the Rohingya community. These posts were inherently harmful and resulting in a direct impact on their rights to equality and non-discrimination.[21]

There is evidence that the key institutional drivers behind the advocacy of hatred against the Rohingya on Facebook proactively and successfully leveraged Meta's content-shaping algorithms to dehumanize and demonize the Rohingya.[22] Viral engagement tactics which are "[t]atics such as the use of clickbait headlines, explicit requests for likes and shares, or the tagging of multiple accounts on posts were regularly used by assets believed to belong these groups' [Myanmar Military/Buddhist national groups] network.[23] These strategies appear to have been used to maximise newsfeed distribution and boost content reach, as well as to build and maintain an engaged audience.

There were multiple direct communications sent to Meta from experts and civil society experts from as early as 2012, and continued every year between 2013 to 2017, warning it of the risk that it could contribute to a genocide or another form of mass violence unless it took urgent action to address issues related to the use of the platform in Myanmar.[24]Amnesty documents in its report of an international aid worker that was a subject of a viral Facebook post that made threats to his life and his difficulty of taking action against the posts against him, despite several complaints being raised on the platform by him.[25]

FAILURE OF META TO MODERATE AND REMOVE HATEFUL CONTENT ON ITS PLATFORM

The content moderation algorithms and measures that Meta employs, whilst useful in picking up some forms of hateful speech and content, do not prevent the algorithmic amplification of the such content.[26] Further, Meta does not invest equitably in content moderation across jurisdictions and languages around the world.[27] Meta employs a tier system which determines the amount of resources it invests in moderating content in specific jurisdictions.[28] Notably, Tier 0 countries, of which there are heightened resources, are countries that are based in the "Global North" or are of notable concern to Meta.

Meta's platforms are used extensively across the globe and are in some jurisdictions "the internet", yet the company fails to adequately invest in content moderation and resources to curb the impact of their products. Whistleblower, Frances Haugen in sharing her testimony before US Congress stated, "[t]he reason why I wanted to do this project is because I think the Global South is in danger". Meta has flagged certain countries such as Myanmar, Ethiopia and Pakistan as high risk, yet still fails to have sufficient content moderators.[29] In the midst of the conflict in Myanmar in 2017, the region only had five Burmese speaking content moderators, and many from the Rohingya community had reported hateful content

---

[19] *Myanmar: The social atrocity, previously cited,* p. 44.
[20] *Myanmar: The social atrocity, previously cited,* p. 13.
[21] United Nations Human Rights Committee, General Comment No. 18, UN Doc. RI/GEN/1/Rev.9 Vol. I (1989), para. 1
[22] *Myanmar: The social atrocity, previously cited,* p. 38.
[23] Victoire Rio, "The Role of Social Media in Fomenting Violence: Myanmar", 2020.
[24] *Myanmar: The social atrocity, previously cited,* p. 32.
[25] *Myanmar: The social atrocity, previously cited,* p. 39.
[26] The Facebook Papers, "We are Responsible for Viral Content", 11 December 2019, p.8- "- https://facebookpapers.com/
[27] "The Tier list: How Facebook decides which countries need protection- https://www.theverge.com/22743753/facebook-tier-list-countries-leaked-documents-content-moderation
[28]"The Tier list: How Facebook decides which countries need protection"- https://www.theverge.com/22743753/facebook-tier-list-countries-leaked-documents-content-moderation
[29] *Myanmar: The social atrocity, previously cited,* p. 32.

CONTEMPORARY FORMS OF HATE SPEECH ONLINE
SUBMISSION TO THE SPECIAL RAPPORTEUR ON CONTEMPORARY FORMS OF RACISM, RACIAL   4
DISCRIMINATION, XENOPHOBIA AND RELATED INTOLERANCE

Amnesty International

"over a 100 times" to no avail.[30] The platform has indicated that it has since made improvements to the platform, but according to a report by Global Witness in March 2022, content moderation systems were still failing to pick up anti-Rohingya and anti-muslim content.[31]

KEY RECOMMENDATIONS:

- <u>For States:</u>
    - o Ban targeted advertising on the basis of invasive tracking practices, such as cross-site tracking and tracking based on sensitive data or other personal data.
    - o Regulate technology companies to ensure that content-shaping algorithms used by online platforms are not based on profiling by default and must require an opt-in instead of an opt- out, with the consent for opting in being freely given, specific, informed and unambiguous.

- <u>For Big Tech Companies such as Meta and Google:</u>
    - o Cease the invasive collection of personal data which undermines the right to privacy and threatens the right to freedom of expression, freedom of thought and other human rights, whilst also contributing towards the fuelling to hateful and harmful speech and content online.
    - o End the practice of using tracing-based advertising and embrace less harmful alternative business models such as contextual advertising.
    - o To protect people's privacy and to give them real choice and control, a profiling-free social media ecosystem should not only be an option but the norm. Therefore, content-shaping algorithms used by online platforms should not be based on profiling by default and must require an opt-in instead of an opt-out, with the consent for opting in being freely given, specific, informed, and unambiguous.
    - o Meta should cooperate fully with the OECD NCP process in the United States brought by the Rohingya, and any other processes that may arise from this complaint, and fully fund the education programming requested by Rohingya communities who are parties to the complaint.

---

[30] *Myanmar: The social atrocity, previously cited,* p. 32.
[31] *Myanmar: The social atrocity, previously cited,* p. 32.

CONTEMPORARY FORMS OF HATE SPEECH ONLINE
SUBMISSION TO THE SPECIAL RAPPORTEUR ON CONTEMPORARY FORMS OF RACISM, RACIAL   5
DISCRIMINATION, XENOPHOBIA AND RELATED INTOLERANCE

Amnesty International