1

WHAT WAS USING FACEBOOK LIKE IN THE LEAD UP TO THE ROHINGYA CRISIS?

Digital Coup

MYANMAR INTERNET PROJECT REPORT

Facebook "substantively contributed to the level of acrimony and dissension and conflict."

MARZUKI DARUSMAN

HEAD OF THE UNITED NATIONS FACT-FINDING MISSION ON MYANMAR

Much has been written about Facebook's role in the Rohingya crisis and its implications for tech accountability. More will undoubtedly follow.

We asked Myanmar digital rights activists - who had a front row seat throughout Facebook's Myanmar history, both as users, living and breathing the platform, and as researchers and advocates, with opportunities to engage the company from as early as 2012 - for their perspective.

WHAT WAS USING FACEBOOK LIKE IN THE LEAD UP TO THE ROHINGYA CRISIS?

Hate speech and disinformation was always a big part of the Facebook experience, as was graphic violence - which would flare up in the days following specific incidents.

When Facebook started becoming popular in Myanmar, in 2011-2012, we were going through a period of political transition and some people feared that change. Ultra-nationalists and some monks were concerned that Myanmar would lose its traditional values. They saw Islam, in particular, as a threat (Myanmar is predominantly Buddhist) and were quite vocal about their fears on Facebook. Facebook enabled them to broadcast their fears to large audiences and made it easy for people to engage with and build on these fears. It also enabled these monks and ultra-nationalists to connect with one another and build communities.

2

WHAT WAS USING FACEBOOK LIKE IN THE LEAD UP TO THE ROHINGYA CRISIS?

Most notably, Facebook was instrumental to the emergence of a mass Buddhist nationalist movement, which grew from 2012 to 2015 to encompass hundreds of thousands of members across the country and came to be known as Ma Ba Tha (Patriotic Association of Myanmar). Ma Ba Tha made extensive use of Facebook, leveraging the platform to build hundreds of local chapters, recruit members, fundraise, organize protests and events and run campaigns. As a movement, Ma Ba Tha launched a campaign to pass a set of four Race and Religion laws, including a monogamy law, a religious conversion law, an interfaith marriage law and a population control law. All four laws were eventually passed in 2015, after a big campaign on Facebook which further stigmatized and vilified Muslims.

With the campaign unfolding months to a general election, the military and the political sphere widely endorsed Ma Ba Tha and its messages. This helped to normalize the anti-Muslim and anti-Rohingya discourse. As the narratives became increasingly politicized, there was also a lot of disinformation, which sought to further stoke fear. Disinformation fueled distrust between communities and triggered a number of violent intercommunal incidents, which made fears of the Muslim community even more real and immediate for people.

By the time the military launched its clearance operations against the Rohingya community in Rakhine State, first in October 2016 and again in September 2017, the public had been primed to fear Muslims, and in particular the Rohingya, who personified the idea of an external threat of Islamisation from the 'Western Gate'. As reports of attacks on military outposts made their way onto Facebook, most people bought the official government narrative, which framed the attacks as terrorism and branded Rohingyas as Muslim terrorists. The weeks that followed the Arakan Rohingya Salvation Army (ARSA) attacks on military outposts, in both years, were particularly dreadful. Facebook went wild with rumors, graphic violence and threats of further attacks. People were afraid and hungry for information. That made them particularly susceptible to sharing unverified claims and to refer back and echo many of the narratives they had previously heard. A lot of out of context images were also used to depict growing horrors - and they were widespread rumors that Muslims were getting ready for larger attacks, and an all out for 'jihad.'

Below are a few examples of the type of content and narratives which were widespread on Facebook both before and as the violence escalated, as well as a brief review of the roles played by different actors.

The meme reads: "don't stop until we kill them all", with the caption calling on readers to give a like or comment if they agree
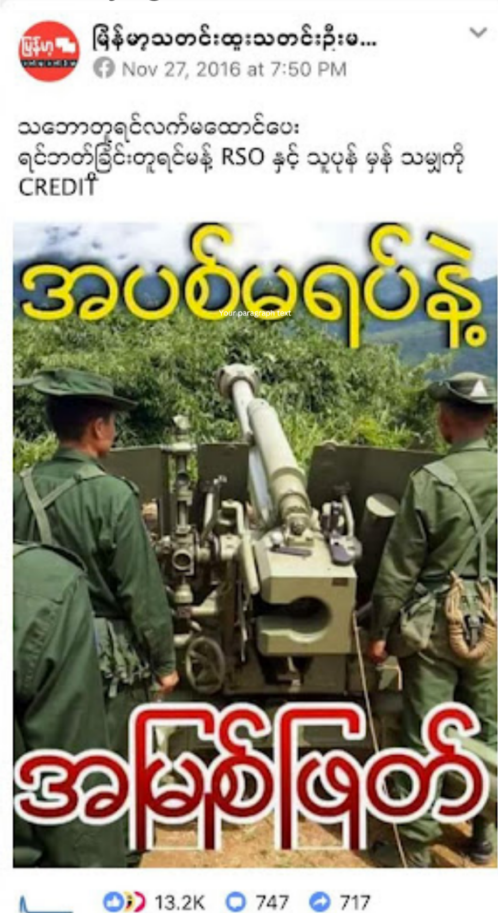
HOW DID THE COMPANY RESPOND TO THE SITUATION - AND YOUR WARNINGS?

We became concerned very early on that social media was a double edged sword.

In 2012 and 2013, Myanmar faced a wave of intercommunal violence. Many people died. In several of the cases, the violence had been triggered by unverified - and often outright fabricated - rumors, which had spread on Facebook.

We tried to warn Facebook at every occasion we had, starting in 2012, that hate speech and disinformation were becoming a serious problem, but they didn't seem to see it as their responsibility and simply ignored us.

This went on for a couple of years until we enlisted the help of a Harvard professor, Susan Benesch, who headed the Dangerous Speech Project and helped us get their attention. Even then, Facebook deflected its responsibility, making the risk situation all about the users: "people lack digital literacy", "speech norms are the problem", "people aren't reporting enough".

We made it clear to them - time and again - that what we were dealing with was organized actors, with a political agenda. It wasn't just about people being naive or lacking literacy. But the few things they did focused on users - translating the reporting system, the settings, the rules, and supporting the occasional awareness raising campaign.

It wasn't until 2018, after they had been formally called out in relation to the Rohingya genocide, that they faced to the fact that their platform was actively being weaponized and started taking action on the actors, both through dangerous actor designations and coordinated inauthentic behaviour takedowns.

Below is an overview of the warnings we gave or know the company received through to 2017 - as well as their response. We'll update this list as we become aware of more formal warnings or actions that may have taken place over that period.

4

WHAT WAS USING FACEBOOK LIKE IN THE LEAD UP TO THE ROHINGYA CRISIS?
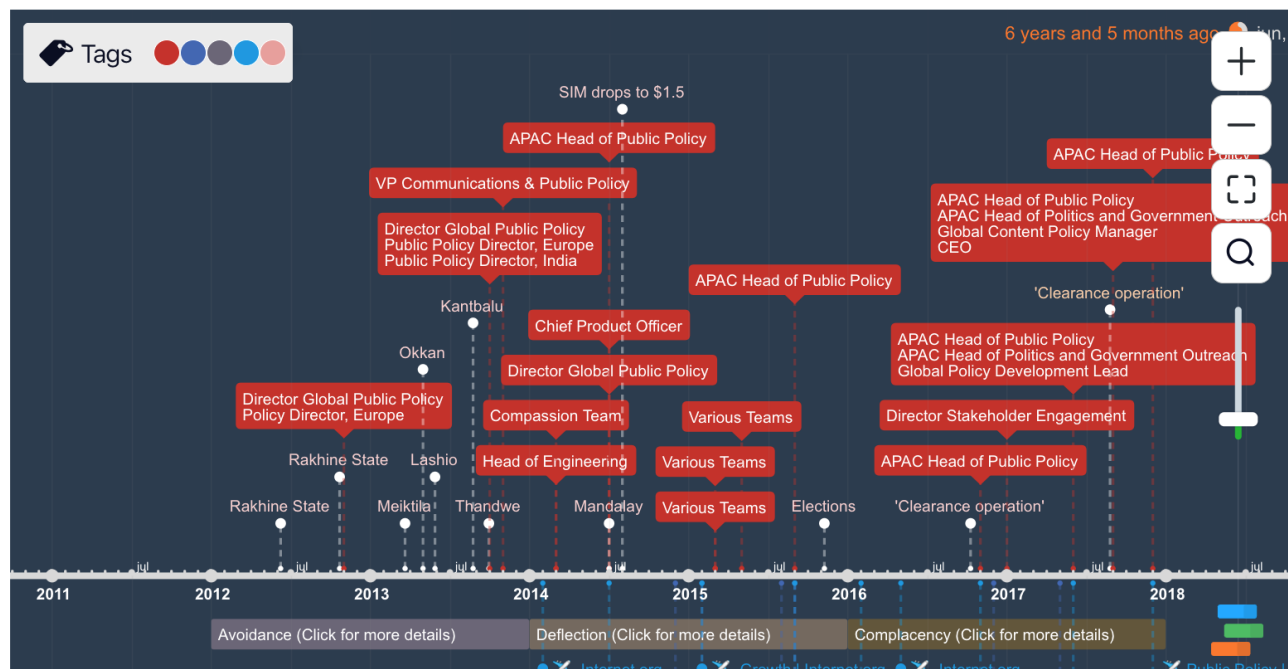
2012 - 2013: Avoidance

The years 2012-2013 were marked by a wave of intercommunal riots and incidents, targeting both Rohingyas and Muslim communities across the country. Facebook was growing quickly and we were getting concerned about the virulent hate speech which was circulating on the platform. We briefed Facebook executives, including Facebook's Director of Global Public Policy, Policy Director , Europe and Policy Director, India, whom Htaike Htaike Aung, co-founder of MIDO, met at international conferences on two distinct occasions - in November 2012 in Baku and in October 2013 in Bali. We also followed up over email to ask for ways to get Facebook to review problematic content and address emergency escalations. They never responded. The platform was getting more and more users, eventually reaching the 1 million mark. The situation was getting so concerning that we started doing targeted research on hate speech. Both the local and international press were starting to pay more attention to the issue. We had warned Facebook. They knew how to reach us. But they didn't.

At the end of 2013, journalist Aela Callan eventually managed to get a meeting with Facebook's VP of Communications and Public Policy, Eliot Schrage, at Facebook's headquarters. He responded to her warnings of inter communal violence and genocide by linking her up with the Facebook compassion team, whose focus at the time was preventing bullying.

2014 - 2015: Deflection

In early 2014 the situation was getting worse. We launched the Panzagar campaign ('flower speech') to rally opposition to hate speech. Together with Aela, Htaike Htaike Aung visited Facebook's Menlo Park headquarters and met with the compassion team. At the same time, working with Matt Schissler, we got connected to a professor at Harvard, Susan Benesh,who arranged for Matt to present on the Myanmar situation to Facebook staff in March 2014, and, together with Matt and Aela, helped bring us into a group call with the compassion team. The compassion team responded to our concerns by offering to localize their tool for reporting objectionable content. They worked with all of us over the course of a few months to refine and translate the reporting flow. By the end of 2014, Facebook had a Burmese language reporting tool - but it was a road to nowhere: they did not have the capacity to review user reports. In May 2014, we were told by Facebook that they only had a lone Burmese language reviewer. According to Reuter, this grew to 2 by February 2015, and eventually to 4 by the time Myanmar went through its general elections, in November 2015. At that point, Facebook had 7.5 million users and abuse, fueled by Ma Ba Tha's race and religion law campaign as well as the elections, was widespread.

The link between online content and offline violence was also becoming increasingly clear by then. A fake rumor of rape, shared by Wira Thu, had triggered major riots in the city of Mandalay in 2014,leaving two dead. This had led the government to shut down Facebook and call in an emergency forum. Facebook's APAC Head of Public Policy, who oversaw Myanmar, took her first official trip to the country on this occasion. She visited a few more times subsequently.

We kept warning Facebook that the problem was one of organized hate that wouldn't get solved just with translated policies (they localized their community standards) or guidance on how to behave online (they commissioned a series of campaign resources and videos). Matt Schissler and David Madden, then CEO of Phandeeyar, each made a trip to Menlo Park in March and May 2015 respectively, to stress the risks and potential consequences for the country. Matt was asked by a Facebook employee if he thought a genocide against the Rohingya was possible. He said yes. Our warnings were clear, but Facebook wasn't listening. Rather than to invest in improving and scaling their enforcement capacity, the company focused its interventions on public facing activities with high PR potential, which deflected its responsibility and shifted the burden for security onto users,

who were regularly portrayed in conversations as either not behaving properly, or not doing enough to report violations.

2016 - 2017: Complacency

By 2016, with the 2015 elections safely behind us, Facebook became complacent. By then, they had Burmese language community standards, a localized reporting tool, a few Burmese language videos on online bullying and hate speech, and direct links to civil society partners who could escalate high risk and emergency issues to their attention. They seemed to think that they had fulfilled their safety requirements towards their Myanmar users - and started to shift their focus towards monetizing their business in the country. In February 2016, Facebook met with Myanmar media and marketing agencies in Yangon, Myanmar, to pitch their publisher program - instant article - which positions Facebook as an intermediary between advertisers and publishers. The program was eventually rolled out in April 2016. In May 2016, Facebook also launched Free Basics and Facebook FLEX in partnership with Myanmar's state owned telecom provider MPT. We were not briefed nor consulted on the risks associated with the introduction of either of these services, and it's unclear that Facebook ever did a human rights impact assessment.

As the situation of the Rohingya quickly deteriorated in October 2016, Facebook was unequipped to proactively address risk concerns. They relied nearly exclusively on us, as local partners, to point them to problematic content. Upon receiving our escalations, which were made over a mix of emails, Facebook Messenger, Facebook Group, WhatsApp or Skype, they would typically address the copy we escalated but take no further steps to remove duplicate copies or address the systemic policy or enforcement gaps that these escalations brought to light. The time it took to address escalations also varied greatly, with evening escalations (most frequent), subject to the greatest delays, as Australia - where the Myanmar market was managed from - was asleep. We kept asking for more points of contact, better escalation protocols, and interlocutors with knowledge of the language and context who could make decisions on the violations without requiring the need for translators and further delays. We got none of that. When we met with the manager in charge of escalations, upon his visit to Yangon in June 2017, we were told that our best bet was to report using the user reporting tool in parallel to our escalations as reports submitted through that tool, he claimed, were being reviewed within an average of 6 hours. That obviously wasn't true. Our own data later showed that the median turnaround time on a report, assuming a review took place, was closer to 48 hours. Facebook had no Burmese language classifiers at the time either, which could have helped triage and prioritize reports.

Despite the escalating risks, we did not see much progress over that period, and Facebook was just as unequipped to deal with the escalation of anti-Rohingya rhetoric and violence in August 2017 as they had been in 2016. They had made some attempt at leveraging automated moderation using slur lists in early 2017, but had run into challenges. Ultimately, it was still down to us, as local partners, to warn them. We simply couldn't cope with the scale. On 9 September 2017, Victoire Rio, then Social Impact Director at Phandeeyar, warned in an email that a Facebook Messenger campaign had the potential to trigger countrywide violence. Her warning couldn't have been more explicit. She barely got a response. In fact, it wasn't until December 2017, when Facebook first visited the country past the events, that she was told that her warnings had made it to Mark Zuckerberg. In spite of the explicit warning, there was no apparent attempt at addressing the broader risk concerns in the months that followed the attacks. Mark Zuckerberg later referred to this period as a success, claiming that Facebook's 'systems' had successfully prevented violence. We called him out in an open letter in April 2018, prompting a formal apology.

6

WHAT WAS USING FACEBOOK LIKE IN THE LEAD UP TO THE ROHINGYA CRISIS?

WOULD YOU SAY THAT FACEBOOK WAS COMPLICIT IN WHAT HAPPENED TO THE ROHINGYA?

Facebook knew that its platform was being weaponized against Muslims in Myanmar, and that the situation of the Rohingya minority was particularly precarious. We warned them. The media warned them. Academia warned them. Myanmar became 'the' textbook example of dangerous speech and social media weaponization. Their own compassion team was looking at the country as a case study. Executives were briefed on it as a prime example.

Not only did they know about the weaponization of their platform in Myanmar, but they also knew the risks. They had seen first hand how Facebook rumors had fueled deadly riots - on multiple occasions. One of their executives had even visited Myanmar after one such incident left two people dead in Mandalay, Myanmar's second biggest city, and after the government, out of options, had resorted to shutting down the platform.

They knew, but they didn't take appropriate measures to address the problem - and we watched as the same risk patterns repeated themselves, over and over again. Their teams knew that the integrity measures they had in place were inadequate and far from on par with the mitigation measures they had in place in the US and other Western countries. They often had a pretty good grasp of what would help too, but the investments didn't follow. In some cases, the investments would have had substantial cost implications, like hiring more content reviewers, which made them redhibitory. In others, however, the required investments would have been minimal, like tweaking a word in the reporting UI, yet even those types of investments were not prioritized.

Not only did Facebook fail in its responsibility to moderate its platform in Myanmar, the company's business model also made the situation materially worse. As a platform, Facebook was not neutral. The company amplified and recommended content which it determined would keep users on its platform longer, as part of its growth strategy. Dangerous content, which triggered strong reactions and sparked debate, were unnaturally amplified by the platform's algorithms. Facebook also offered distribution as a service, and accepted money from the very actors it knew were weaponizing the platform, allowing them to reach both large and targeted audiences. As it sought to expand its Burmese language ads real estate, the company also launched its instant article program in Myanmar. The program was launched without adequate safeguards, and fueled the emergence of a new type of content publishers which thrived on sensationalistic and clickbait content. This fundamentally altered and sensationalized the Myanmar information environment, and played an important role in the mainstreaming of anti-Rohingya narratives, as well as in fueling fear in the wake of the attacks.

Below are a few concrete examples of security vulnerabilities which we flagged to Facebook and which were not addressed at the time.

7

WHAT WAS USING FACEBOOK LIKE IN THE LEAD UP TO THE ROHINGYA CRISIS?

Current 'photo view' UI prompts users to 'report post'. Until mid 2020, the UI used to misleadingly prompt users to 'report post' - when only the photo would get through the review process
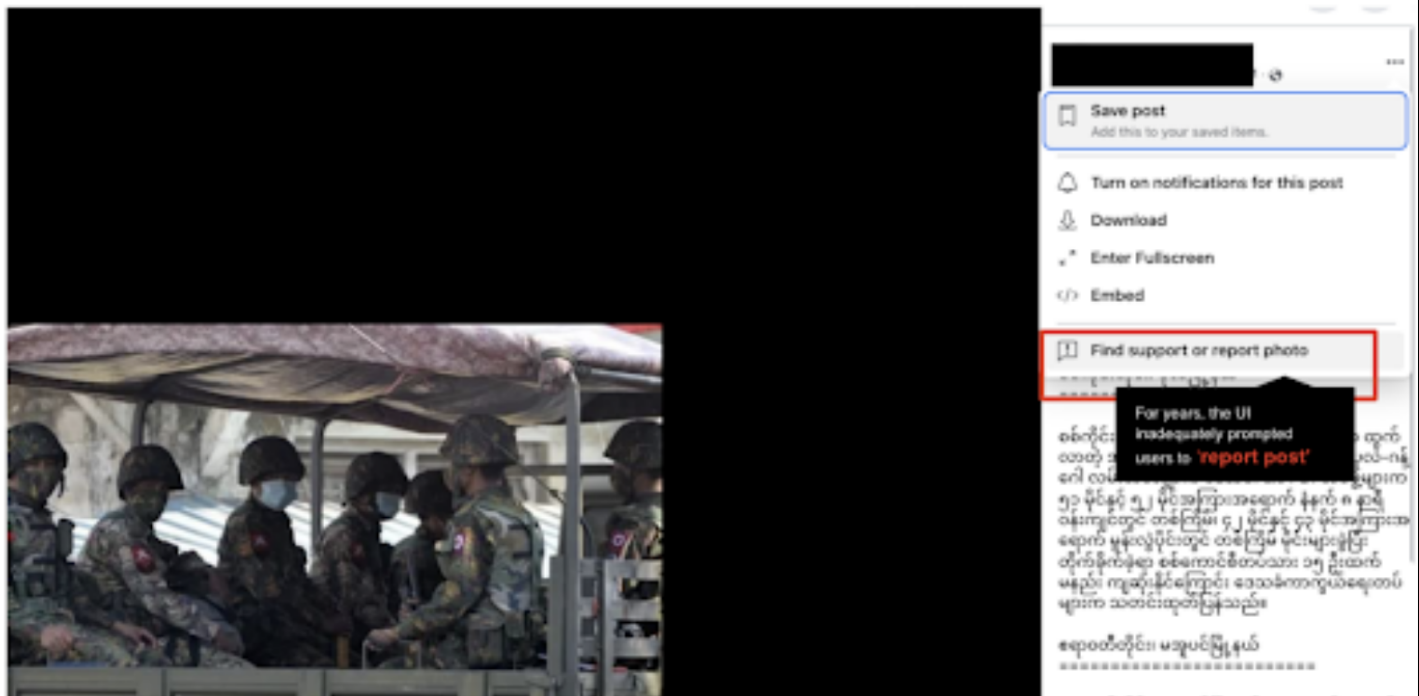
TO WHAT EXTENT DO YOU FEEL THAT THE PROBLEMS HAVE NOW BEEN ADDRESSED?

Facebook was called out very publicly in 2018 for their role in the Rohingya crisis. The UN fact finding mission called them out in their report. We called them out in an open Letter. We also worked with Senator Leahy and Senator Blumenthal to raise questions to Mark Zuckerberg, when he was questioned by the US Congress. Ultimately, Facebook had to admit that they had not reacted quickly enough and done enough in Myanmar. They committed to doing more and have ramped up their investments and engagement in the country since.

Facebook's experience in Myanmar triggered some important changes, which carry global significance and are worth stressing. Most notably, Facebook finally acknowledged that context matters, and endorsed the need for contextualized policies to account for specific risk environments. This is an important departure from their prior standpoint, which emphasized the global nature of policies, and prevented Facebook from acting on a number of

8

WHAT WAS USING FACEBOOK LIKE IN THE LEAD UP TO THE ROHINGYA CRISIS?

serious escalations we made for fear of disproportionately impacting on freedom of expression in other markets. This prioritization of context was also positive in that it drove the company to hire Myanmar expertise and to step up its engagement with local civil society. Another significant development was the company's acknowledgement that it needed to do more to restrict access to the actors responsible for weaponizing its services. They designated a number of actors over the weaponization of the platform in the Rohingya crisis, including military officials and entities, covert assets attributed to the military, and Ma Ba Tha and affiliated spokespersons. Though enforcement remains a challenge (see below), the Myanmar experience provides a clear case for deplatforming, with the designation of Ma Ba Tha, in particular, being critical to bringing the movement into check and mitigating its ability to commit further harms.

These developments are important progress but they are not enough. Much more needs to be done if we are to prevent a repeat of what happened to the Rohingya and ensure that Facebook, and other similar platforms, do not contribute to exposing communities to heightened security risks, the world over.
As a priority, we need platforms to:

1. Invest in enforcement

    Facebook's enforcement capacity is still inadequate in Myanmar. Violating content still regularly goes unmoderated, which in turn results in violating actors not facing strikes or sanctions for their recurring abuse. We also continue to deal with undue removals, often explained as AI errors. These are problematic as they tend to disproportionately impact on activists and the media who are most vocal and political on the platform. The situation is even worse with other platforms, such as YouTube - whose enforcement capacity for Myanmar is still minimal, or Telegram - which is just ignoring all reports of violation. It is not enough for platforms to have policies that describe what is prohibited. They need to invest in their capacity to enforce these policies. Facebook and other companies need to increase their pool of language and context-aware human reviewers; Where leveraging slur lists, they need to ensure that these are properly contextualized and subjected to regular audits and independent reviews; They also need to prioritize developing AI classifiers for minority languages and ensure that these are regularly audited to guarantee accuracy and prevent overenforcement.

2. Scale actor-level moderation

    While Facebook did acknowledge its responsibility to restrict access to actors with a track record of weaponizing its services, and nominally restricted several individuals and entities involved in the anti-Rohingya campaign, it continues to be unable to reliably prevent these actors from using its services. As a matter of fact, most, if not all, of the key stakeholders in the anti-Rohingya campaign continue to maintain a presence on Facebook and to leverage Facebook and other platforms for influence. As we repetitively warned the platforms, the bulk of the harmful content we face comes from a handful of actors, who have been consistently violating Terms of Services and Community Standards. Facebook and other companies need to rethink their moderation approach to more effectively deter and - where warranted - restrict actors with a track record of violating their rules and terms of services, including by enforcing sanctions and restrictions at an actor and not account level, and by developing better strategies to detect and remove accounts of actors under bans.

3. Make amplification an earned right

    Though Facebook introduced a number of measures to try to prevent the amplification of harmful content in Myanmar, with some success, disinformation, in particular, continues to regularly reach audiences in the millions. According to disclosures from disinformation publishers with a well established presence in

Myanmar, upward of 90% of their views, as of 2022, come from Facebook's algorithmic amplification, rather than their own audience or the shares received on their content. This is true even for pages which were recently established, with some disinformation pages managing to get more reach than Myanmar's most established independent media within days of operation. Facebook and other companies need to commit to making algorithmic amplification an earned right, which requires content authors to meet certain trustworthiness criteria, rather than rely primarily on signals based on a content's actual and predicted engagement. Companies should be transparent about their trust criteria and undertake regular audits of their recommendation algorithms to ensure that they do not amplify problematic content.

4. Be responsible - and accountable - for who they do business with

Facebook not only took money from the military and Ma Ba Tha, in exchange for helping them distribute their content to wide audiences. It's also been channeling funds, through its monetization programs, to a range of publishers spreading disinformation and hate for profit. By failing to do due diligence on these actors, Facebook subsidized the creation of a new industry, which is becoming an increasing problem in Myanmar and elsewhere. Facebook and other companies need to commit to not funding hate and disinformation actors. They need to conduct adequate due diligence on who they sell ads to and on behalf of. Monetization program partners should be publicly disclosed and a mechanism to report harmful actors provided. Monetization program review and appeal processes should also be regularly audited.

While it's critical that platforms take more responsibility and accountability for mitigating the harm of their products and services in *all* of their markets, we also want to stress our strong disagreement with existing calls for platforms such as Facebook to get out of markets like Myanmar. Countries like Myanmar should not be written out of global platforms. To even suggest it is deeply unhelpful and has the potential to be incredibly harmful. Big Tech platforms are accessed and used globally. Restricting access to a handful of countries where the human rights situation is troubling is not going to prevent people from using these services. It will just make it harder for people to exercise their freedom of expression and have the counter-productive effect of providing legal cover for platforms to do nothing to protect users in those countries

---

Recommended Citation Style –

Myanmar Internet Project (2022, September). https://www.myanmarinternet.info/rohingya-crisis .

MYANMAR
INTERNET PROJECT

မြန်မာအင်တာနက်အစီအစဉ်ကို ဒီဂျစ်တယ်နယ်ပယ်၌ နှစ်ပေါင်းများစွာ အတွေ့အကြုံရှိသောသုတေသနသမား များ၊ ကျွမ်းကျင်သူများ၊ ဒီဂျစ်တယ်လွတ်လပ်ခွင့်မြှင့်တင် လိုသူများနှင့် တက်ကြွလှုပ်ရှားသူများဖြင့် ၂၀၂၂ ခုနှစ်တွင် စတင်တည်ထောင်ခဲ့သည်။

မြန်မာအင်တာနက်အစီအစဉ်မှာ ပူးပေါင်းပါဝင်လိုပါ သလား။

မြန်မာအင်တာနက်အစီအစဉ်ကို စတင်၍ အကောင်အထည်ဖော်နေပြီဖြစ်ပါသဖြင့် ဒီဂျစ်တယ် ရပ်ဝန်းမှ စိတ်ဝင်စားသူများ ပူးပေါင်းပါဝင်လာမှုကို ကြိုဆိုပါသည်။ အောက်ဖော်ပြပါ အခြေအနေများ အတွက် မိတ်ဆွေ၏ ကူညီပံ့ပိုးမှု လိုအပ်ပါသည်။

- မြန်မာအင်တာနက်အစီအစဉ်တွင် မရှိသေးသော သုတေသနစာတမ်းများ။
- အလွတ်တန်းသုတေသနသမားများ၏ ထုတ်ဝေဖြန့် ချီခြင်း မပြုရသေးသော သုတေသနစာတမ်းများ။
- အများပြည်သူအကျိုးအတွက် ကောင်းမွန်စွာ အသုံးပြုနိုင်သော အချက်အလက်၊ မှတ်တမ်းနှင့် ကိန်းဂဏန်းများ။
- အင်္ဂလိပ်ဘာသာစကား အသုံးပြုသူများထံသို့ မိမိတို့ ၏ စာတမ်းများ၊ ထုတ်ဝေစာများနှင့် အချက်အလက် များ ကျယ်ပြန့်စွာ ရောက်ရှိလိုသူများ။

ဆက်သွယ်ရန်
info@myanmarinternet.info
www.myanmarinternet.info

Layout   Unknown

Design Concept Simply Me

Photos   ©MIP