Toda Peace Institute

Policy Brief No. 78 June 2020

The Role of Social Media in Fomenting Violence: Myanmar

Victoire Rio

Introduction

Myanmar was originally portrayed as an internet success story. Connected in a matter of months after five decades of military dictatorship, Myanmar people were supposed to be able to communicate freely, access unimpeded information online, experience the benefits of a range of compelling and empowering digital services, and connect with the world. While social media opened up the country's information and communication ecosystem in an unprecedented way, it also came with its share of challenges. Low levels of digital and media literacy, amidst a context of political transition, dominated by rising inter-communal tensions and the formalisation of a Buddhist nationalist movement, made for a dangerous cocktail. Facebook, which came to dominate the Myanmar internet ecosystem, failed to acknowledge and effectively mitigate the risks, letting its platform get abused and weaponised. This case study looks at the role of Facebook in fomenting violence in Myanmar. It explores how the social media platform was weaponised, reviews Facebook's response to the situation, and draws learnings and recommendations for how risks may be better mitigated, moving forward.

The Political Context: An Emerging and Fragile Democracy

A country fresh out of a military dictatorship

Myanmar, previously known as Burma, was under a military dictatorship for nearly five decades. The country embarked on a democratic transition process in 2011. The National League for Democracy (NLD), led by Nobel prize winner Aung San Suu Kyi, boycotted the 2010 election which brought to the presidency a former military general. In November 2015, the NLD won a landslide victory, taking effective control of the government in 2016.

Complex power sharing arrangement

Although the election of the NLD was a move away from military rule, the military continues to exert significant influence over Myanmar politics to this day. In line with the 2008 constitution, the military retains 25% of parliamentary seats as well as control of three key ministries (Defence, Home Affairs and Border Affairs), giving it direct authority over Myanmar's armed forces, police and border guards. Former military officers also continue to occupy positions of authority across all levels of government, the judiciary and many state-owned enterprises.

Diverse population

Myanmar has a population of over 53 million. Its population is ethnically diverse, accounting for the country's geographic location at the intersection of South, Southeast and East Asia. Though predominantly Buddhist, Myanmar is also religiously diverse, with Christian, Muslim, Hindu and animist communities found across the country.

The military regime split Myanmar's identity into eight major ethnic races, themselves broken down into 135 "national races". The Rohingya, a Muslim minority from northern Rakhine State, are notably excluded from this list.

Rising intercommunal tensions

Although not new, the controversy around the citizenship rights of the Rohingya gained renewed focus as Myanmar engaged in its political transition. It was further intensified as tensions between ethnic Rakhine and Rohingya flared up following the alleged rape and murder of a Buddhist Rakhine woman in May 2012, and as attacks by the Arakan Rohingya Salvation Army (ARSA) in 2016 and 2017 triggered military clearance operations in the region, sending nearly a million Rohingya to flee to Bangladesh.

As the Rohingya issue merged into a broader anti-Muslim narrative, tensions between Buddhist and Muslim communities have also escalated countrywide, including in communities where Buddhists and Muslims used to live in harmony. Deadly altercations took place in Meiktila (2013), West Bago Region (2013), Mandalay (2014) and Lashio (2014), and the country has been affected more broadly by violent mobs targeting individuals and property, as well as limitations on the religious freedom and freedom of movement of Muslim minorities.

Formalisation of an organised Buddhist Nationalist movement

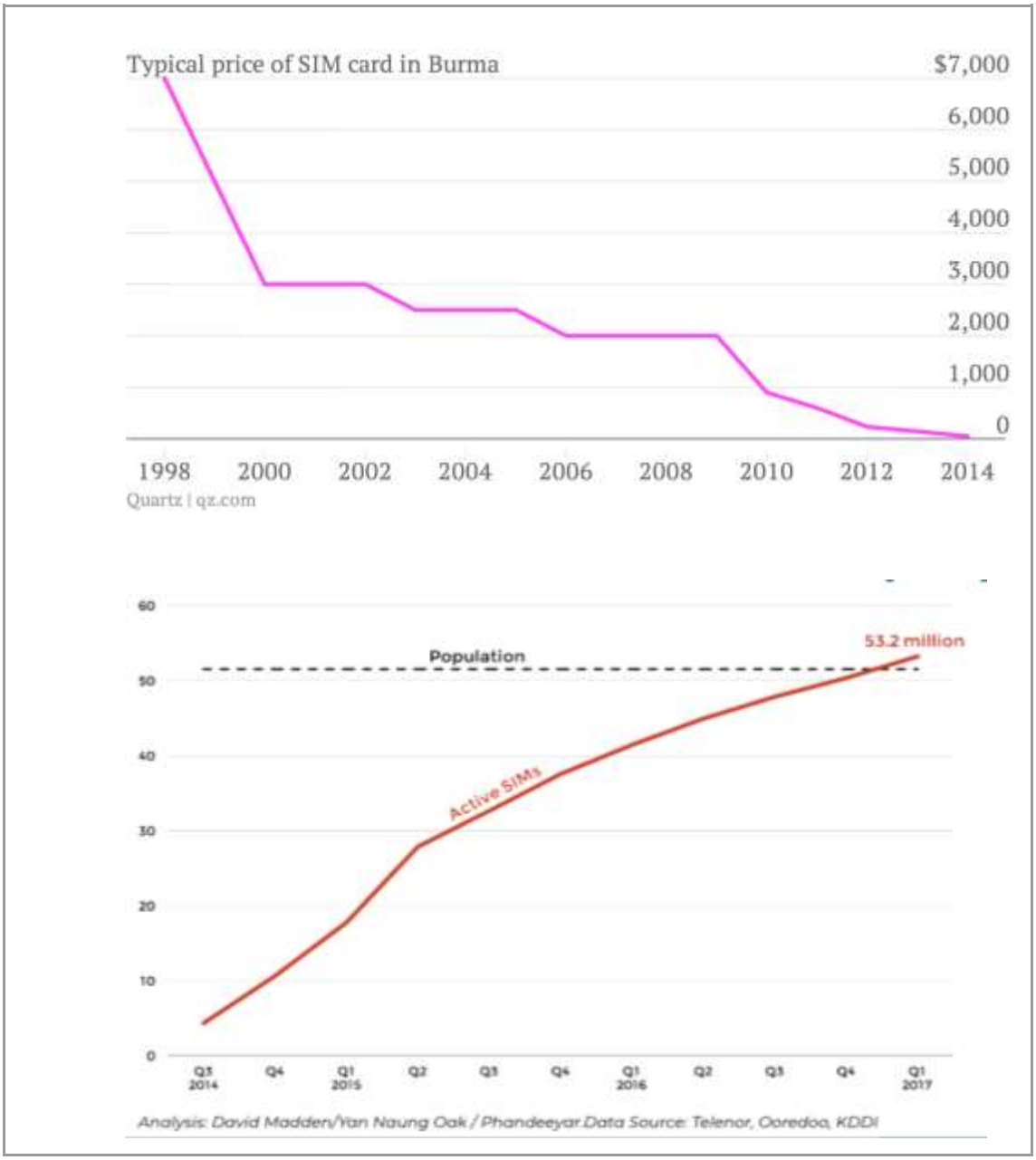
The rise in inter-communal tensions was fueled by, and fed into, the formalisation of an organised Buddhist nationalist movement, with some high-profile monks leading their following into political debates. The 969 movement emerged as a response to what was seen as a rising and existential Muslim threat to Buddhism – driven by an internal threat linked to inter-religious marriage and high birth rates, and an external threat linked to immigration and terrorism. Banned in 2013 by the Buddhist clergy, 969 paved the way for Ma Ba Tha (the Organization for the Protection of Race and Religion), a more organised network with a significant local presence at the township level and a member base reportedly exceeding 100,000. Ma Ba Tha achieved political prominence by advocating for a set of Race and Religion Laws, which were passed in 2015 following an intense campaign that built on a stigmatising and occasionally dehumanising anti-Muslim narrative. Though formally banned in 2017, Ma Ba Tha has continued to reinvent itself under different branding and their ideology continues to exert influence in many circles across Myanmar.

The Internet Landscape: Facebook First

A connectivity revolution

During its military years, Myanmar was largely disconnected from the outside world. Public information was heavily controlled and digital access severely constrained by poor infrastructure and high prices. In 2013, the country started liberalising its telecommunications sector, resulting in the introduction of two foreign telecommunication operators (Telenor and Ooredoo) in 2014. The new operators immediately drove prices down – from the equivalent of two hundred US dollars for a SIM card in early 2014 (more than the average household monthly expenditure) to approximately $1.50 USD, significantly increasing affordability.

As prices dropped and infrastructure projects multiplied, millions of people across the country gained access to smartphones and mobile data. Nowhere has the connectivity revolution been so drastic. Myanmar, a country of 53 million, went from having an internet penetration of 0.23% in 2011 to 39% in 2019 with the number of active SIM cards surpassing people by 2017.



Facebook as the de facto internet

Facebook grew in popularity as more and more people accessed mobile data and it came to form the centre of Myanmar internet users' experience, becoming a near synonym for the internet.

Even before connectivity became widely affordable, in 2014, Facebook had hundreds of thousands of existing users, many of whom were accessing the platform through devices belonging to friends or family members. Gaining their own uninterrupted access to Facebook was a key driver behind the rush for smartphone ownership. Smartphones came pre-installed with the Facebook app, with phone shop clerks offering to set up multiple accounts and passwords for users for a small fee.

Myanmar people were hungry for entertainment and information and quickly became avid Facebook users. Media and entertainment companies responded by investing heavily in digital content production. Services, which would normally be found on websites or apps, also started leveraging Facebook, turning the platform into a one stop shop for finding hotels, restaurants, booking bus tickets, engaging in online shopping and even booking doctor appointments. As most services and most content in the various Myanmar languages resided on Facebook, the broader web did not gain much traction – a phenomenon further explained by the fact that most Myanmar users had little or no experience with web browsers.

Facebook's growing dominance of Myanmar's information ecosystem was further supported by the offering of Free Basics, the free version of Facebook, from 2016 to 2017, and the integration of Facebook and Messenger, which quickly became a favourite for instant communication.

So important is Facebook to the Myanmar information ecosystem that the news of the resignation of President Htin Kyaw, in March 2018, was first broken on the Myanmar President Office Facebook page and his successor's election organised as a Facebook event by the parliament.

Myanmar President Office first publishes the statement of resignation of President Htin Kyaw on Facebook on 11 March 2018 at 10:41am

Union Parliament of Myanmar sets up an event for the election of the new president on 28 March 2018.

Digital Novices

Myanmar users did not grow accustomed to technology as it developed. From a controlled information environment, they went straight into information overload, with little or no time to develop the critical digital and media literacy skills needed to serve as a foundation for internet safety. The virtual and seemingly anonymous nature of the web further provided users with a false sense of safety and freedom, which led many to see their online identity as detached from their offline identity. This altered their sense of responsibility and self-restraint, normalised the practice of "friending" strangers, and created a sense of detachment from one's personal account. This in turn normalised the notion that one could have more than one account (two accounts are considered normal), while also reducing the importance and relevance of privacy. It is not uncommon to see Myanmar people set up new accounts whenever they get logged out or share passwords or even accounts with others – a practice particularly favoured by young couples, who see sharing accounts or account details as a display of trust and love. Such practices have made Myanmar users particularly vulnerable to scams and harassment, while also creating a favourable climate in which irresponsible speech and misinformation prosper.

The Weaponisation of Facebook

The military and Ma Ba Tha—the Buddhist nationalist organisation—were amongst the first to recognise the potential of Facebook in Myanmar, which they went on to use for propaganda, surveillance and coordination purposes.

Strategic and coordinated leveraging of Facebook

The military, which has been sending officers to study in Russia since 2001[1], was among the first adopters of the Facebook platform in Myanmar. It was quick to launch a dedicated curriculum as part of its Defense Service Academy Information Warfare training and reportedly set up specialised social media units dedicated to producing and promoting content, engaging in social media monitoring, and trolling critics. According to the New York Times, these units included as many as 700 officers as of October 2018, working in shifts in troll farms.

Ma Ba Tha monks were similarly amongst Facebook's early adopters, seeing in the platform a major way to support the development of their movement. In a 2016 interview with Buzzfeed, Wirathu, one of the leading monks behind the Ma Ba Tha movement and a Facebook early adopter, acknowledged that "If the internet had not come to [Myanmar], not many people would know my opinion and messages like now," adding that he had always written books and delivered sermons but that the "internet is a faster way to spread the messages". Wirathu quickly learned that consistency was key to engagement and was

[1] Myoe, Maung Aung. 2006. *Building the Tatmadaw: Myanmar Armed Forces Since 1948*, Singapore: Institute for Southeast Asian Studies,

prompt to establish a social media team to oversee the network social media presence. In addition to counting a dedicated social media committee, Ma Ba Tha is known to have provided regular training to its members on content creation and Facebook features as well as leveraging secret Facebook groups to coordinate their support.

Propaganda

Myanmar has a long history of propaganda – an art which the military and nationalist groups have honed over many years through offline medium, such as fliers, pamphlets, magazines, books, CDs and DVDs, as well as events and sermons. What Facebook offered these groups was unprecedented possibilities for reach, targeting and speed, which they were able to access at little to no cost and use to gain an oversized influence on the public discourse.

Reach: Although Facebook access was originally limited to the cities, the arrival of new telecommunication operators quickly brought access to more rural areas. At the time of the August 2017 attacks, Facebook had approximately 18 million active users across the country. Whilst significant for a country of 53 million, these numbers should be seen in the light of Myanmar's broader information dissemination dynamics, where people actively share information with family and friends.

Targeting: Not only was the potential for reach significant, but Facebook also offered distinct targeting advantages. It facilitated the cheap and rapid testing of messages, allowed for the gathering of insights on people's interests, tastes and aspirations, and offered possibilities for micro targeting, both organically through the creation of targeted communities, as well as through the use of paid ads.

Speed: Facebook further offered unprecedented speed, allowing for real time publishing and making it possible for these groups to promptly react and build on real life events, which allowed them to capitalise on the raw emotions these incidents generate.

*Little to no cost***:** Compared to offline materials, which require large financial investments to produce and distribute, Facebook offered a free medium for content dissemination, as well as a cheap option for targeted ads.

Though the military and Ma Ba Tha may have had only a few hundred staff working on their social media strategy full time, these groups were able to exert an oversized influence on the public discourse by relying on widespread networks of Facebook assets combined with strong marketing savviness.

Widespread networks of assets

Both groups relied on widespread networks of assets to distribute and amplify their propaganda:

*Official accounts, pages and groups***:** Both the military and Ma Ba Tha maintained an active official presence on the platform, until Facebook banned their respective use in 2018.

Most leaders had their own accounts, giving them access to up to 5,000 'friends' and hundreds of thousands of 'followers'. In addition to running active accounts, both groups also used institutional pages and groups to carry their message. By the end of 2017, the official page of Commander-in-Chief Senior General Min Aung Hlaing had over 1.3 million followers, while his office page counted over 2.7 million. Wirathu—one of Ma Ba Tha's most well-known monks—achieved a following of over 400,000. Despite being a well-known hate figure, and being sanctioned several times, Wirathu was able to find his way back to the platform on multiple separate occasions, before he was permanently banned in January 2018[2].

Senior General Min Aung Hlaing official page	
Commander-in-Chief's office page	

Fake accounts and 'alternative' presence: The military and Ma Ba Tha also operated covert presences, allowing them to amplify their messaging, and engage in covert activity. This included the use of fake accounts, as well as 'alternative' pages and groups – which though not acknowledging their affiliation, engaged in the repeated and systematic sharing of the groups' content and messaging.

A New York Times investigation found, for example, that the military was behind a large number of entertainment pages, which would include misinformation and propaganda amongst otherwise trivial content, such as celebrity gossip or health and beauty tips. Facebook confirmed the existence of such covert operations, taking down a total of five hundred and ninety three pages, thirty two groups and two hundred and twenty six accounts as part of four successive coordinated inauthentic behaviour takedowns in

[2] Wirthu was charged with sedition in 2019 and has been on the run ever since.

2018 and 2019. Pages impersonating real media and public figures are also widely believed to make up part of the military's distribution channels.

Influencers: In addition to managing overt and covert assets, both groups appear to have benefited from a network of influencers, including prominent writers and singers, who further contributed to the amplification of their messaging and agenda.

Leveraging of amplification

Both groups also demonstrated strong marketing savvy, which enabled them to amplify their messages beyond their own networks:

Significant capacity for content production: Large amounts of time and resources appear to have been dedicated to the production of content, with new content introduced daily. While some of the content explicitly advanced the groups' agenda or was designed to manipulate popular opinion through the spread of rumours and disinformation, other content appeared to be intent on maximising attention and audience building. Memes, cartoons, blogs and videos, designed for virality, were regularly shared, showing a high level of understanding for the social media demand and high technical skills. Content from various creators, including regular Facebook users, was also widely repurposed, with little consideration for copyrights or credits.

One of the most sophisticated endeavours from Ma Ba Tha included the production of a six-minute video re-enacting the rape and murder of Thi Dar Htwaye, the Rakhine woman whose death kickstarted the 2012 wave of violence in Rakhine state. According to the Myanmar Times, the video, which was first posted through Wirathu's account in early 2016, generated over 8,000 shares and over 120,000 views before it was taken down for violating Facebook's community standards. Particularly successful content (including this video) was regularly repurposed and reuploaded, generating additional reach every time.

Use of viral engagement tactics: In addition to demonstrating capacity and skills for content production, both groups showed a keen understanding of the Facebook platform and viral engagement strategies. Tactics such as the use of clickbait headlines, explicit requests for likes and shares, or the tagging of multiple accounts on posts were regularly used by assets believed to belong to these groups' networks. These strategies appear to have been used to maximise newsfeed distribution and boost content reach, as well as to build and maintain an engaged audience.

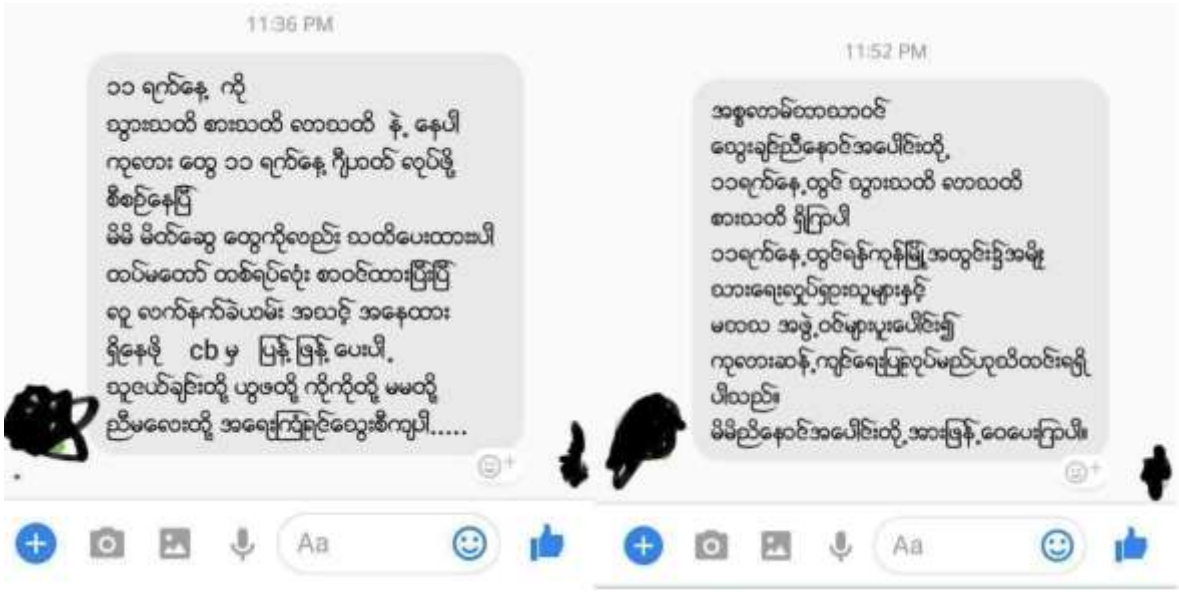
Audience segmentation: Facebook's page and group features were also used to develop and engage audiences, as well as enable tailored message distribution. Township level groups, for example, appear to have been used by both the military and Ma Ba Tha to support the targeting of hyper local communities.

Use of Facebook Ads: Besides using pages and groups to target their audience, both the military and Ma Ba Tha have been found to make use of Facebook's boosting and ad features, which allow for the micro-targeting of audience by location, age, gender and a range of interests and behaviour patterns.

Examples of sponsored content by official pages from Ma Ba Tha and the Military

Leveraging of Facebook Messenger: In addition to leveraging the Facebook main platform, civil society believes that the military experimented with Facebook messenger as part of its effort to spread fear and incite violence. The most famous example of such use was the circulation of parallel rumours of imminent attacks in September 2017, which appear to have been designed to ignite violence between Muslim and Buddhist communities while the military crackdown was underway in Northern Rakhine. These rumours were shared widely across the country, by Facebook accounts often unknown to those receiving the message, and were amplified by extensive reference across various accounts, pages and groups, which sought to give them credibility.

Mass messaging campaign circulated on Facebook Messenger between 6-11 September 2017



Rough translation:

On the right: “Be warned and stay alert. The Kalar are planning to launch a Jihad on Monday 11 Sept. Warn your friends. The order to get ready with guns has already been issued in the army. Please forward this message in the chat box. Friends, brothers and sisters, let’s unite."

On the left: "Dear Islam brotherhood, be warned and stay alert. On Monday 11 Sept in Yangon, Ma Ba Tha and extremist nationalists will collaborate and they will launch an anti kalar movement. Please forward this message to our brothers.

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

Surveillance

Besides providing a platform for widespread propaganda campaigns, Facebook provided a unique tool to monitor, discredit and attack critics and opposing views. Aside from the very real physical and mental risks to the targets, this helped create a climate of self-censorship, which further enabled the military and Ma Ba Tha to dominate the narrative and shape public opinion.

Monitoring: Facebook provides a unique window into the life, thoughts, activities and interests of its users. Though search features were progressively restricted following global privacy concerns, both groups are believed to have made use of the platform to map, monitor and target critics.

Trolling: Trolls, widely believed to be affiliated with the Military or Ma Ba Tha, were also regularly found commenting on high visibility pages and groups. They often worked in "brigades", taking over the comment section of a particular post and working in coordination to discredit any criticism and/or attack the critic.

Mass reporting: Besides discrediting content through trolling, military and Ma Ba Tha trolls are also widely believed to have engaged in mass reporting campaigns against their critics, in an effort to get their content or account banned by Facebook.

Doxxing: Another strategy used to deter critical posting has included publishing misinformation against vocal critics along with their private information. Such posts have included information such as the critic's profile link, their photo, their address, their ID number or their religion. Some posts have gone as far as to include identifiable information about family members.

Direct Messaging: A number of activists and journalists have also reported being threatened over direct messages, sometimes by multiple accounts at once.

Hacking: Besides being verbally abused and harassed, critics have also been the target of hacking attacks, designed to either discredit them, reveal their personal information, or gain access to their network.

Legal threat: Some critics have also been the victims of legal attacks, leveraging different articles within Myanmar law that criminalise defamation. One of the most notable examples is the case against the editor of Myanmar Now, an independent investigative media, who was sued by a Ma Ba Tha sympathiser for a Facebook post which put into question Wirathu's monk status.

Coordination

Beyond providing a platform for propaganda, and a tool for surveillance and targeted attacks, Facebook provided the backbone infrastructure which helped coordinate these groups' activities, and facilitate their outreach and—in the case of Ma Ba Tha—their fundraising:

Messenger: Facebook messenger is a critical means of communication in Myanmar. It is used as a prime alternative to traditional phone lines, SMS and emails and is widely believed to have been used to support communication between group members through both direct messages and large messenger threads.

Secret Groups: Secret Facebook Groups are also believed to have played a central role, particularly in Ma Ba Tha's coordination effort, enabling it to coordinate its activities, its

trolling and its mass reporting campaigns as well as to store and exchange content quickly and easily.

Facebook Events: Both Ma Ba Tha and the military have also made extensive use of Facebook's event feature to organise rallies, demonstrations and protests. Besides helping them to promote events broadly, the feature is believed to have enabled these groups to track attendance, build their audience and further refine their targeting.

Facebook Live: Ma Ba Tha has also been found to leverage the Facebook live feature as a way to broadcast and generate a visual (and shareable) archive of its events.

Friend-ing: Finally, it is widely believed that both groups have capitalised on Myanmar users' tendency to accept friend requests from people they do not know as part of their monitoring efforts as well as to engage potential new followers and supporters.

A sense of impunity

The widespread and sustained weaponisation of the Facebook platform in Myanmar was made possible by Facebook's 'hands off' approach to moderation in the country. Sanctions were rare and, when applied, mostly addressed individual pieces of content which, by the time they were removed or demoted (48hrs +), had already achieved their intended impact.

In rare cases where accounts or pages were disabled as a result of an investigation or repeated community standards violations, actors were able to find their way back to the platform. In most cases, they were able to do so using the same name and photo, which allowed them to quickly rebuild their following. Wirathu, who made the cover of Time magazine as "The Face of Buddhist Terror" in 2013 and was officially banned from preaching in 2017 for his widespread use of hate speech, for example, was able to create at least three successive profiles on the platform, with his last known official account only banned in January 2018.

Facebook's Response

Absentee landlord

Facebook has been referred to as an "absentee landlord" in the context of Myanmar. Despite repeated warnings around the abuse and weaponisation of its platform, starting in 2012, the company did not visit the country until 2014. It first visited on the invitation of the Myanmar government, which invited Facebook to participate in a roundtable discussion following deadly riots in Mandalay which had been triggered by a Facebook rumour. According to Wired, the government sought Facebook's support during the situation but was not able to make contact and proceeded with shutting down the platform altogether. From this visit onward, the company adopted a bi-annual visit schedule, with the country managed from Australia – a situation which lasted through to 2018, when the company acknowledged that it should do more and started visiting more frequently.

Lack of contextual understanding

Besides lacking the necessary time and resources to engage with the Myanmar situation, the team in charge of Myanmar lacked the language skills and context to effectively keep up with local developments. Concerns over the company's lack of understanding of the context and rapidly deteriorating situation were repeatedly raised by Myanmar civil society organisations, including through at least five in-person briefings in Menlo Park beginning in 2014. Despite the repeated warning and urges to hire local expertise, Facebook did not hire Myanmar staff until early 2019.

Meagre Human Rights due diligence

While the UN Guiding principles on Business and Human Rights place a responsibility on companies to uphold human rights, it is unclear to what extent Facebook engaged in the necessary Human Rights due diligence when rolling out its product in Myanmar. The first sign of such an effort being undertaken was the commissioning of an independent Human Rights Impact Assessment, which was published in November 2018, four years after the platform acquired its first million users in the country and more than five years after receiving its first warnings of abuse and weaponisation.

Limited use of AI and automated detection

While Facebook claims to be leveraging technology to proactively detect community standards violations, it is unclear to what extent the company leveraged automation as part of its Burmese language moderation prior to 2018.

As per the company's own reporting (here and here), only 13% of the content removed in the 4th quarter of 2017 was proactively identified. Judging by the number of reported removals in Myanmar in the 3rd quarter of 2018 (64,000), this would have represented at most 2000-3000 self-identified violations a month, raising questions about the scale and effectiveness of the system.

The mass takedown of content containing the keyword 'kalar' in May 2017 (a derogatory slur used against Muslims of South Asian descent – but also a composite in harmless words such as kalar pae /chickpea or kalar htine / chair) also raised widespread concerns over the use of overly simplistic keyword-based detection, failing to account for linguistic or contextual considerations and negatively impacting on freedom of expression.

Insufficient investment into human moderation

Not only did Facebook have minimal capacity for automated detection in Burmese language, it also failed to adequately invest in human moderators. According to Reuters, Facebook hired its first moderator for Burmese language content in 2013, a year after concerns over misuse of the platform first made it into the US media. The company then scaled its moderation effort from two to four Burmese content moderators in 2015 to cope with the heightened risks associated with the historic elections taking place that year. Despite the rise in real world violence following the October 2016 and August 2017 attacks in Rakhine

state, the company had only "dozens" of reviewers moderating Burmese content by the time Mark Zuckerberg was called to testify in front of the US congress in April 2018. Judging by the company's recruitment ads at the time, it is also unclear how much linguistic and contextual understanding these reviewers had. The position of Myanmar Market Specialist, for example, had Burmese language skills listed as an 'ideal' rather than an 'essential' requirement.

Meanwhile, outsourced reviewer jobs with companies like CPL, one of Facebook's contractors at the time, required applicants to be legally authorised to work in Ireland, which severely restricted the pool of candidates and introduced a clear bias towards members of the Myanmar diaspora, who had often left Myanmar while the country was still a military regime and before Facebook became widely available and used.

Lack of effective emergency mechanism

With Facebook lacking the capacity to self-moderate its platform, it relied extensively on escalations from third parties to address dangerous content and emergencies. In the absence of a formal escalation mechanism, however, these escalations had to be channeled via email or messaging apps to policy staff, who lacked language and context, were not empowered to take action, and worked outside of Myanmar working hours. As a result, emergencies, even when flagged, were often not addressed quickly enough to mitigate their impact.

Focus on individual content appeals and PR campaigns rather than structural enforcement solutions

Working through the policy team also presented major limitations as policy staff were neither trained nor empowered to identify and address systemic flaws with the product and its inner workings. As a result, most of Facebook's response to the deteriorating situation in Myanmar focused on developing, or commissioning, communication materials and on addressing case by case appeals. These solutions placed the bulk of the responsibility on Myanmar users and civil society and did little to address the root causes of the problem, which lay in the systematic abuse and weaponisation of the platform by coordinated actors.

A belated awakening

In April 2018, after being called out by the UN, Myanmar civil society, the US congress and the US media, Facebook CEO Mark Zuckerberg acknowledged that Facebook had been too slow to respond to the abuse and weaponisation of its platform in Myanmar. He pledged to dedicate more resources as well as to establish a cross functional team to mitigate the risks of Facebook contributing to real world harm in Myanmar.

In the months since, Facebook has ramped up its efforts in the country, including hiring a dedicated Myanmar public policy team, increasing content reviewer capacity, rolling out Burmese language AI classifiers and engaging far more widely with local civil society.

Recommendations

Whilst Facebook's failure to effectively mitigate risks has had particularly drastic consequences in Myanmar, the Myanmar experience with the platform is not unlike what many other countries across the global south experienced with Facebook, as well as other similar platforms. The following section draws on the experience of Myanmar to derive actionable lessons and recommendations for social media companies and civil society:

To Social Media Companies:

1. Prioritise human rights due diligence

Facebook commissioned a Human Rights Impact Assessments (HRIA) for Myanmar in 2018 – more than five years after the company was first informed that its platform was being actively weaponised and only after Facebook had been widely criticised for facilitating a campaign of ethnic cleansing. Though the recommendations were helpful, they came much too late and nearly one and a half years on, are still to be fully implemented.

- Social Media companies should commission country-level HRIAs, regardless of whether they have a physical presence in a country, as soon as they become available in a market – and at the latest before they become mainstream. These assessments should consider how the company may be contributing to adverse human rights impact, not only through its product(s) but also through its staffing and operations. These should be conducted by an independent third party, involve consultation with local civil society and rights holders, and be released publicly, in their integrality. At a minimum, HRIAs should be renewed every five years, unless the human rights situation evolves or the company is made aware of widespread human rights concerns linked to its product(s) or operations.
- Social Media companies should further develop country-level human rights risk mitigation strategies. These should identify specific risk mitigation measures for each of the risks identified in the HRIA, and lay out clear targets, timelines and responsibilities for operationalisation. The implementation of these strategies should be overseen by a human rights executive, reporting directly to the Company's senior management team, and strategies should be reviewed and updated regularly as the context evolves.

2. Engage directly with local civil society

Since 2018, Facebook has dramatically scaled up its engagement with Myanmar civil society. In addition to undertaking a number of country visits, which have allowed for the direct exposure of different product and research teams, the company has established a dedicated Myanmar team and organised a series of consultations with stakeholders across the country. Unlike their global counterparts, local civil society actors have direct experience with the product and the way it is used in their market. This makes them uniquely positioned to provide detailed and granular feedback. These engagements have gone a long way towards helping Facebook develop a more sophisticated understanding of the Myanmar context and should serve as a model for other countries.

- Social Media companies should make every effort to consult local experts and civil society directly on matters of relevance to their country. Wherever possible, local teams should be established to enable in-person and local language engagement.
- Social media companies should increase exposure of their product teams to different user realities, and where possible, facilitate direct engagement with civil society in countries facing high risk of human rights abuse.
- Social Media companies should also conduct adequate due diligence to ensure that high risk users and minorities are adequately represented in consultations and that their safety is not being compromised by their engagement. In countries where the context and situation warrant it, companies should be prepared to provide direct channels to their regional or global staff to ensure engagement of minority groups which may not feel comfortable engaging with national staff.
- Social Media companies should also commit to following clear and adequate engagement practices. In particular, they should clearly distinguish between information sharing and consultation, and where consulting, communicate the purpose and scope of the consultation as well as provide the consulted party with sufficient notice and background information to allow them to shape an informed opinion.

3. Contextualise policy using a tiered approach to protected categories

The experience of Myanmar demonstrates how hate speech targeted at vulnerable minorities can pave the way for widespread human rights abuse. While Facebook's intention to apply the same community standards to all of its users is laudable, Myanmar should remind us that not all "protected characteristics" face an equal threat of harm, and that threat can vary greatly across socio-political contexts, as well as over time. The threat faced by Rohingya Muslims in Myanmar in 2016 and 2017, for example, cannot be equated to the threat faced by White Christians in a North American context or even Rohingya Muslims residing outside of Myanmar.

- Social Media companies should consider applying a tiered approach to their use of "protected characteristics", similar to the way Facebook introduced a tiered approach to its hate speech definition and enforcement. By using different tiers, companies could maintain global policy standards, while still accounting for differences in the threat levels experienced by different "protected characteristics" across contexts and time. The designation of high risk "Protected Characteristics" could support the broadening of restriction to dangerous speech which may otherwise fall under a "grey area", as well as trigger additional resourcing towards detection and moderation. Escalations in status should be temporary, however, and subject to a clear designation protocol and objective criteria, informed by the companies' country-level human rights impact assessment and mitigation strategies.

4. Scale, regularly audit, and continuously improve country-level enforcement

Policies need to go hand in hand with enforcement if they are to make a difference. Since 2018, Facebook has worked hard to scale its enforcement capacity in Myanmar, hiring additional content reviewers as well as rolling out AI classifiers for Burmese language. Though this has made the company more proactive in its moderation, a number of major

challenges remain. Independent research by the author, for example, identified various flaws with Facebook's report triage system, which accounts for systematic and widespread content moderation errors on user reports. This is a serious issue as user reports remain the only path to content moderation for Myanmar's minority users – as well as for a significant number of users globally, whose languages are yet to leverage AI classifiers. Though AI classifiers are now contributing to the detection of potentially violating content in Burmese language, the accuracy of these classifiers also continues to be widely inadequate, resulting in undue sanctions being regularly applied, at times with significant implications for freedom of expression and information.

- Social Media companies should conduct regular market level audits of their content moderation and enforcement processes. Particular attention should be given to reporting flows and triaging systems, AI classifiers, and the recruitment and activity of human reviewers. These audits should be conducted by an independent third party, possibly overseen by an independent oversight body, such as the newly established Facebook Oversight Board.
- Social Media companies should also contribute funding and data to support the development of ethnic language corpora and consider developing scholarships in Natural Language Processing (NLP) and AI targeted at minority and ethnic language communities.

5. Scale efforts to detect and curtail abuse at its source

Though increased capacity for content level moderation is important, it is not enough to deter and root out coordinated abuse. Bad actors spend months—if not years—building networks of online assets, including accounts, pages and groups, that allow them to manipulate the conversation. These inauthentic presences continue to present a major risk in places like Myanmar and are responsible for the overwhelming majority of problematic content.

- Social media companies should invest significantly more resources towards the detection and investigation of coordinated inauthentic behaviour. They should drastically expand their investigation teams and prioritise the development of detection systems, which can help trigger such investigations.
- Social media companies should consider expanding their security bounty model to encompass the detection of information operations and new means of weaponising their platform.

6. Preserve evidence of abuse

Although the existence of coordinated efforts to weaponise the Facebook platform for political ends in Myanmar is now widely accepted, the ability of independent researchers and accountability bodies to understand the full scope and scale of the operations is heavily constrained by the continuous and ongoing removal of assets and content by Facebook and the actors themselves.

- Social Media companies should consider retaining all content removed for policy violations for a period of at least two years.
- Social Media companies should further consider making available to researchers and accountability bodies all public assets such as accounts, pages, and groups identified

as being part of a coordinated inauthentic behaviour operation – including the content which had been previously removed for other policy violations.

To Civil Society:

1. Standardise, systematise and consolidate evidence collection

When presented with evidence of abuse in Myanmar from 2013 through early 2018, Facebook often dismissed concerns as anecdotal and defaulted to one-off fixes. The more civil society was able to consolidate evidence and show patterns, however, the more they became able to push for systemic solutions.

- Wherever possible, civil society should standardise, systematise and consolidate their evidence collection. As a general rule, if a problem seems to be widespread or recurring, civil society should consider complementing qualitative approaches with quantitative ones. There is a growing body of work on methodologies and ways to structure data collection, which can be built on to help shape such research.
- Civil society should also work together across countries to identify shared patterns as well as differences in treatment across markets and engage in joint advocacy.

2. Focus on enforcement and not just policy

A lot of civil society criticism of social media to date has focused on questions of policy and bias, i.e. Where should the line between hate speech and freedom of expression fall? To what extent are Facebook's decision driven by its moderators? Though these are definitely areas that require attention, it is important for civil society to also pay attention to how platforms such as Facebook enforce their policies and whether their enforcement capacity is where it should. Civil society's experience in Myanmar has shown that enforcement capacity varies significantly from country to country and language to language. For languages where Facebook has no text-based AI classifiers, for example, enforcement relies almost exclusively on user and civil society reports to the company. That is a huge limitation. There can also be flaws in the triaging of these reports and not always sufficient language-specific human capacity to review them, which can result in widespread enforcement errors.

- Civil society should advocate for improvements to enforcement, both in terms of resourcing and accuracy.
- Civil society should also keep in mind that social media platforms, at their core, are little more than a combination of databases, workflows and algorithms. As with any system, bugs can and do happen, particularly as changes keep getting introduced. Civil society should develop a basic technical understanding of the systems to be able to actively probe them and study errors which may be rooted in technical issues.

The Author

Victoire Rio is a researcher and digital rights advocate with a background in conflict, peacebuilding, data and technology. She has been living in Myanmar since 2015, where she researches the social media landscape and advises a network of Myanmar Civil Society organisations on advocacy to tech companies and how to mitigate the risk of social media induced violence and political instability.

Toda Peace Institute

The **Toda Peace Institute** is an independent, nonpartisan institute committed to advancing a more just and peaceful world through policy-oriented peace research and practice. The Institute commissions evidence-based research, convenes multi-track and multi-disciplinary problem-solving workshops and seminars, and promotes dialogue across ethnic, cultural, religious and political divides. It catalyses practical, policy-oriented conversations between theoretical experts, practitioners, policymakers and civil society leaders in order to discern innovative and creative solutions to the major problems confronting the world in the twenty-first century (see www.toda.org for more information).

Contact Us
Toda Peace Institute
Samon Eleven Bldg. 5th Floor
3-1 Samon-cho, Shinjuku-ku, Tokyo 160-0017, Japan
Email: contact@toda.org