On Facebook's New Misinformation Policy for Sri Lanka

20 July 2018

Colombo,
Sri Lanka.

Recent news reports have surfaced noting that Facebook will be proactively taking steps to remove misinformation that could lead to people being physically harmed.

Though long overdue, we welcome this announcement and look forward to learning more from the company. Facebook now seems to acknowledge the impact misinformation can have in certain contexts, especially in countries like Sri Lanka, as noted by CEO Mark Zuckerberg himself in an interview with Recode.

Subsequently, a report in the *New York Times* on July 18 noted that this policy of removing misinformation is already in effect in Sri Lanka with plans to roll out in Myanmar and other countries like India. Even though Facebook has interacted with local civil society organisations in the months since riots in Digana, news that the new policies around misinformation were already in effect took us by surprise. No one we are aware of was informed of the new policies, or enlisted as trusted partners to help Facebook combat misinformation on the lines the news reports and the Recode interview suggests.

Accordingly, while Zuckerberg's commitment to address the harm brought about by misinformation in contexts like Sri Lanka and Myanmar is both vital and welcome, we wish to place on record several concerns with the outlines of the policy as reported in the *New York Times*.

> *"Under the new rules, Facebook said it would create partnerships with local civil society groups to identify misinformation for removal. The new rules are already being put in effect in Sri Lanka, and Ms. Lyons said the company hoped to soon introduce them in Myanmar, then expand elsewhere."*

The solution here appears to place the onus on the partner civil society organisations, which we feel is unsustainable in the long-term. Although Facebook's Community Standards are publicly available, their test for determining real-world harm is unclear. It is also unclear whether this test has taken into account the complex contexts underlying the production of content in countries like Sri Lanka or Myanmar, including the varied and nuanced use of language and imagery.

The need for offline violence and harm to be demonstrable in order for content to be taken down is impractical, and especially so in an emergency situation. In the case of an attack on minorities in Sri Lanka and Myanmar, this implies that the content would only be taken down after the fact, when photos or reports of arson and death are provided. Through our monitoring of social media during the Kandy riots In March 2018, we noted that the role of Facebook (including WhatsApp) is more or less complete by the time the actual violence takes place, and addressing it only after an incident has happened is not only futile but irresponsible.

Facebook has committed to improving the language capacity of its moderators, especially in Sinhala. Recent reports have also flagged the proliferation of hate speech and misinformation on the platform in Tamil, which is of particular concern given heightened tensions between Muslims and Tamils in the

East. Yet, this content has not been removed even when reported using the in-app mechanisms over mobile.

Getting this right matters, for users of Facebook in Sri Lanka and for the company. The policies around misinformation and how it is addressed as piloted in Sri Lanka and Myanmar will deeply inform and influence the way false news is tackled on the platform in other countries with a history of sectarian violence and conflict. It is therefore of paramount importance that the solutions implemented are sustainable, scalable, responsive, and replicable.

Until the company's investments in AI and machine learning come online a few years hence, we reiterate that increasing the number and improving the language capacity of moderators looking at content in countries like Sri Lanka and Myanmar is the best solution to address misinformation, and is in fact what Facebook has committed to do in the recent past.