PRODUCT POLICY FORUM

facebook

NOVEMBER 5, 2019

- RECOMMENDATION: Unverifiable Rumors

Recommendation: Unverifiable Rumors

Content Distribution Policy

Issue

We remove content when we know it is false and has the potential to contribute to imminent physical harm or violence. We implement this policy based on input from local partners. We also want to stop the spread of rumors that are likely to contribute to violence even if we can't always functionally verify whether the claims are false. However, we don't want to restrict speech that turns out to be true.

Unverifiable Rumors

Overview

Recommendation: Adopt specific guidelines to define functionally unverifiable claims

External Outreach: 49 External Engagements

Working Groups: 5 Working Groups

Unverifiable Rumors

Status Quo

We remove misinformation that could contribute to imminent violence or physical harm.

- To operationalize the policy, we have on-boarded local organizations and international institutions, who are often the first to become aware of inaccurate or misleading information that may contribute to physical harm or violence.
- We also have secure channels for these organizations to reach out to us and escalate content quickly.

Unverifiable Rumors
Status Quo - Examples

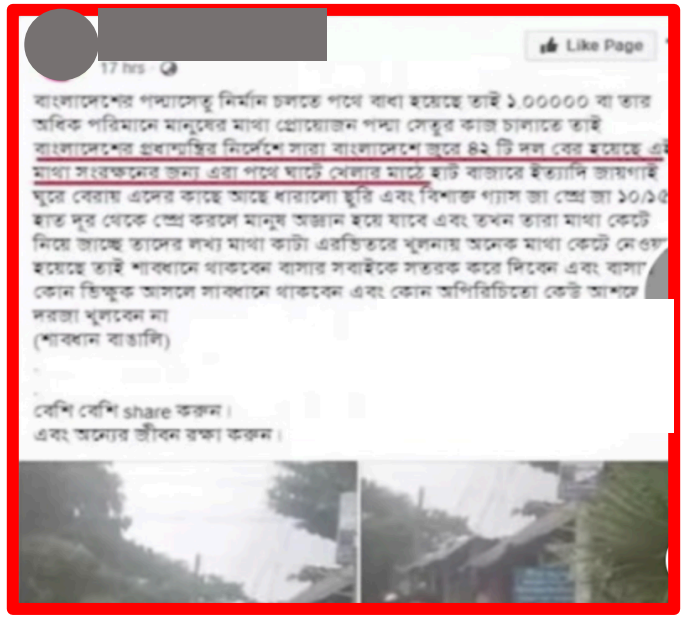
The first two examples do not meet the requirements for removal as misinformation that is false and has the potential to contribute to imminent physical harm or violence because they are not provably false. As unverifiable rumors, however, they still have the potential to contribute to offline harm or violence and are reflective of the type of content we want to capture as we work to update our policy.

Remove (SOP) ✗

"පල්ලි වල ආයුධ ගොඩගහනවා කියලත් දන්නෙ නැද්ද..කැම වලට වද පෙති දානවා දැනගෙන හිටියෙම නැද්ද..ඒවා කලේ අපි අතරම හිටපු උඹලගෙම මිනිස්සු නෙවෙයිද..පාරෙ ත්යෙන කුකුල් මස් කඩ ..වඩේ කරත්ත වල ඒවා කරන කොට අපිත් දැනගෙන හිටියෙ නෑ උඹලා ත්‍රස්තවාදයට සහය දෙන එවුන් බව"

"Didn't you know that they were stocking weapons in the mosques, that they were adding sterilization pills to food..."
(Sri Lanka)

Remove (SOP) ✗



"The construction of a bridge across Bangladesh Padma River halted as it requires more than 100,000 severed heads. According to the statement of the Prime Minister of Bangladesh there are 42 groups who have started their expedition to collect the severed heads..."

Allow ✓

Unverifiable Rumors

Findings From Research and External Outreach

- Most 'harmful' rumors targeted at vulnerable groups (rather than specific individuals) are unverifiable because they lack specificity or make sweeping claims about marginal groups.
- Greatest risk is from rumors with misleading content that have the potential to encourage a violent response against some target.
- Rumors that have led to violence in the past often reoccur, which makes it possible to identify unverifiable rumors with real risks of harm.
- Policy should focus on limiting amplification of high-risk rumors and consider parallel work with trusted local organizations in high-risk settings.
- Likelihood of harm is more important than whether the rumor is true or false.
- Document specific rumors and rumors that may recur.
- Remove, reduce, and inform depending on the type of rumors.

Policy Relevance: Policy should focus on limiting amplification of high-risk rumors and consider parallel work with trusted local organizations in high-risk settings.

Sources: Internal Research; Brass, Paul R. (2011); De Figueiredo, R., & Weingast, B. (1999); Donohue, W. A. (2017); Horowitz, Donald. (2001); Javed, J., & Miller, B. (2019); Oh, O., Gupta, P., Agrawal, M., & Rao, H. R. (2018); Tsesis, A. (2002).; Varshney, Ashutosh. (2003); Metaxas & Mustafaraj (2002); Vosoughi, Roy, Aral (2018)

Unverifiable Rumors

External Outreach

We spoke to 49 experts globally, including academics, human rights experts and civil society orgs.

Unverifiable Rumors

Option 1 (Rec)

Remove claims that are functionally unverifiable and are likely to contribute to violence or physical harm.

Pros

- Flexibility to address inciting rumors that could contribute to violence
- Allows Facebook to remove against a broad variety of rumors

Cons

- Risk of stifling speech
- Risk of removing true content

Unverifiable Rumors
Option 1 - Examples

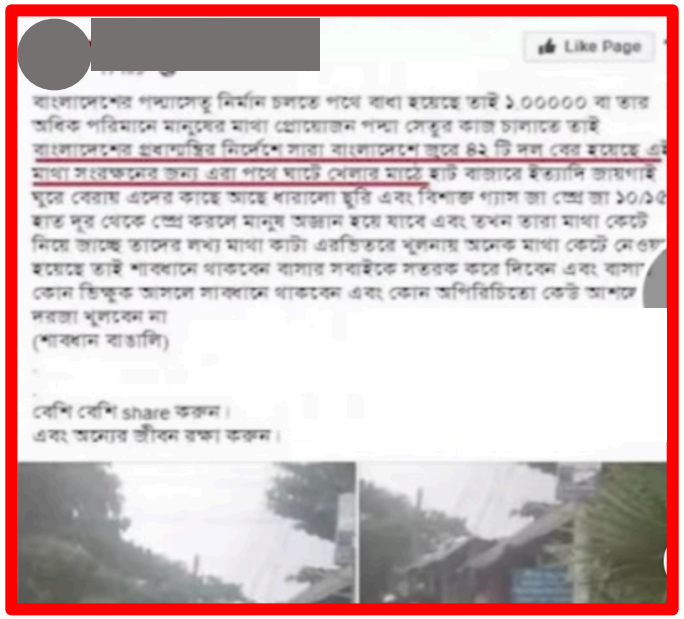
Under the policy proposal we've recommended, the first two examples are removed as claims that are functionally unverifiable and are likely to contribute to violence or physical harm. When needed, we also work with local partners to establish the likelihood of harm.

Remove ✗

"පල්ලි වල ආයුධ ගොඩගහනවා කියලත් දන්නෙ නැද්ද..කැම වලට වද පෙති දානවා දැනගෙන හිටියෙම නැද්ද..ඒවා කලේ අපි අතරම හිටපු උඹලගෙම මිනිස්සු නෙවෙයිද..පාරෙ ත්යෙන කුකුල් මස් කඩ ..වගේ කරත්ත වල ඒවා කරන කොට අපිත් දැනගෙන හිටියෙ නෑ උඹලා ත්‍රස්තවාදයට සහය දෙන එවුන් බව"

"Didn't you know that they were stocking weapons in the mosques, that they were adding sterilization pills to food..."
(Sri Lanka)

Remove ✗



"The construction of a bridge across Bangladesh Padma River halted as it requires more than 100,000 severed heads. According to the statement of the Prime Minister of Bangladesh there are 42 groups who have started their expedition to collect the severed heads..."

Allow ✓

Unverifiable Rumors

Option 2

Remove content that meets the criteria set forth by Option 1 + require that the subject of the claim be a group that shares a protected characteristic.

Pros

- Addresses what many experts consider the most harmful rumors: ethnic, racial, and religious-identity related rumors
- Lower risk of stifling speech

Cons

- Limits Facebook's ability to address non-protected-characteristic rumors
- Risk of removing true content

Unverifiable Rumors

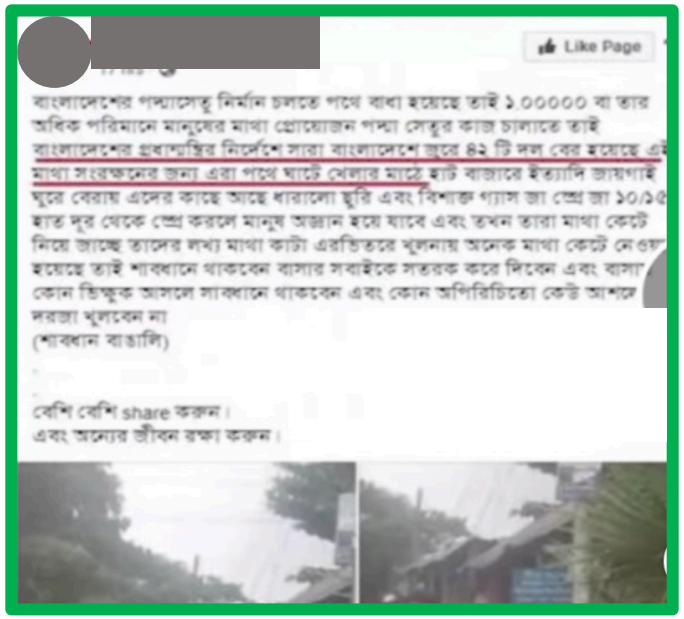
Option 2 - Examples

Remove

"පල්ලි වල ආයුධ ගොඩගහනවා කියලත් දන්නෙ නැද්ද..කැම වලට වද පෙති දානවා දැනගෙන හිටියෙම නැද්ද..ඒවා කලේ අපි අතරම හිටපු උඹලගෙම මිනිස්සු නෙවෙයිද..පාරෙ ත්යෙන කුකුල් මස් කඩ ..වගේ කරත්ත වල ඒවා කරන කොට අපිත් දැනගෙන හිටියෙ නෑ උඹලා ත්‍රස්තවාදයට සහය දෙන එවුන් බව"

"Didn't you know that they were stocking weapons in the mosques, that they were adding sterilization pills to food..."
(Sri Lanka)

Allow

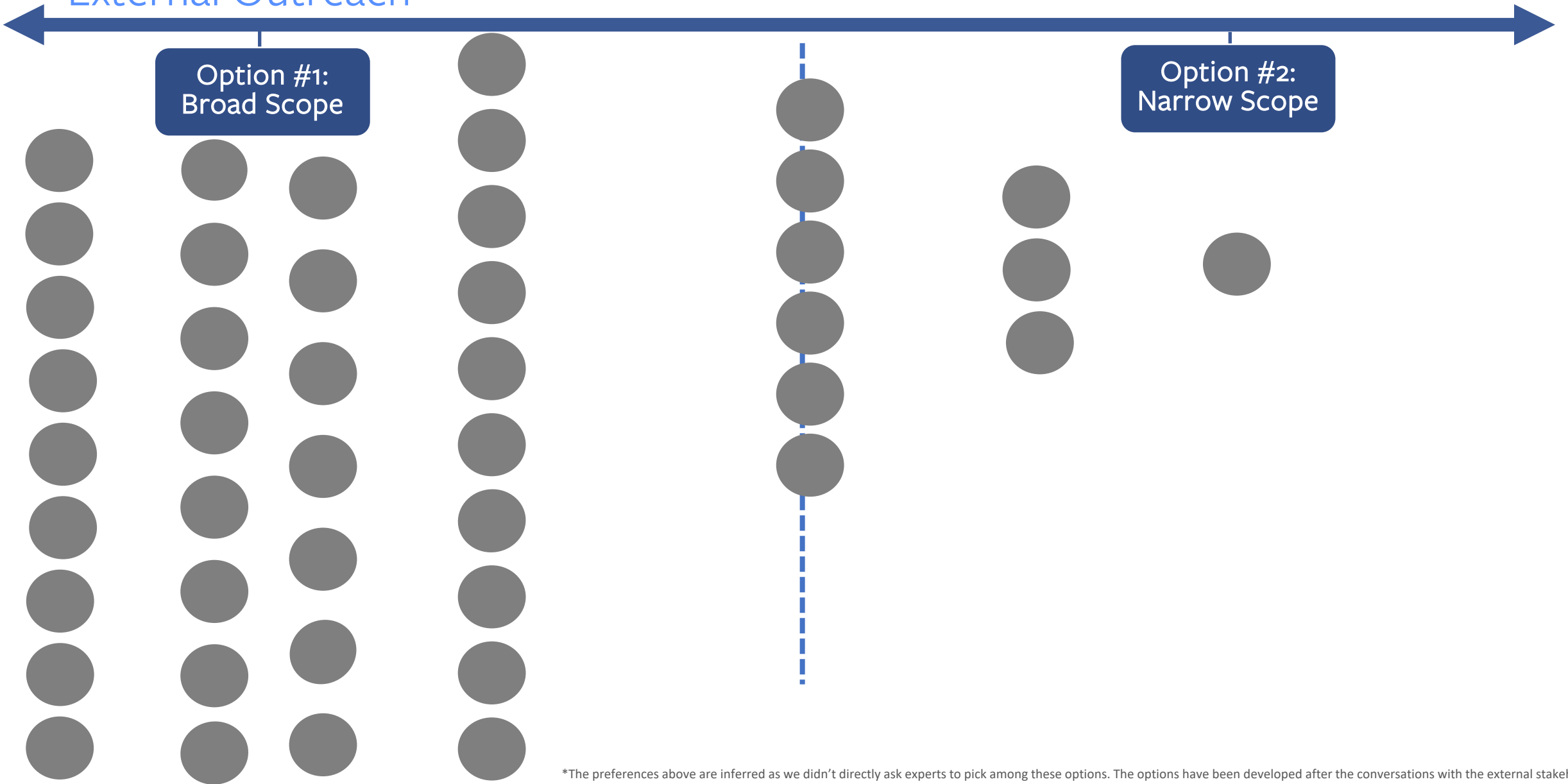


"The construction of a bridge across Bangladesh Padma River halted as it requires more than 100,000 severed heads. According to the statement of the Prime Minister of Bangladesh there are 42 groups who have started their expedition to collect the severed heads..."

Allow

Unverifiable Rumors

External Outreach



*The preferences above are inferred as we didn't directly ask experts to pick among these options. The options have been developed after the conversations with the external stakeholders.

Next Steps

Developments

Reconvene cross-functional working group to decide on enforcement

Comms

Update guidance to local partners

facebook