POLICY FORUM

facebook

June 10, 2020

Agenda

- Recommendation: Evaluation Framework for Outing-Risk Groups

Recommendation: Evaluation Framework for Outing-Risk Groups

Organic Content Policy

Issue

Our violence and incitement policy does not allow people to "out" members of specific high-risk groups, but we do not have a consistent process for identifying high-risk groups or deciding exactly how they should be protected. We want to ensure we apply consistent protections to individuals at risk of harm from being outed; however, we do not want to infringe on discussion and debate of public individuals or important issues.

Evaluation Framework for Outing-Risk Groups
Overview

Recommendation
Our recommendation today consists of two parts:

1. **Evaluation Process**
   A process for assessing whether members of certain groups should receive protection from outing

2. **Proposed Policy**
   Protect the identity and location of members of outing-risk groups and their family members, and protect some professional activities

Source
Recognized inconsistencies in designation of at-risk groups and their protections

External Outreach
31 External Engagements

Working Groups
8 XFN Working Groups

Evaluation Framework for Outing-Risk Groups
Status Quo

- Our Violence and Incitement policy prohibits content created for the express purpose of outing someone as a member of a designated member of a group at risk of outing.

- The current designations process is not standardized across various regions – and neither is the scope of protections offered.

- For example:
    - In Pakistan, we prohibit outing of individuals (but not family members), groups, and locations as Ahmadi, Ahmadiya, or Qadiyani.
    - Allegations of religious conversion, apostasy, or blasphemy against private individuals in Indonesia, Israel, Pakistan, Bangladesh, Maldives, Nigeria, Somalia, Saudi Arabia, Palestine, Egypt, Bahrain, Iraq, Iran, Jordan, Kuwait, Lebanon, Oman, Sudan, Syria, United Arab Emirates, Yemen, Algeria, Libya, Mauritania, Morocco, Tunisia, and Western Sahara.

Evaluation Framework for Outing-Risk Groups
Research Findings

- Risk and vulnerability have various definitions across fields of study and have evolved throughout time.

- External literature review suggests four relevant risk factors in defining vulnerability: legal discrimination, societal and cultural discrimination, practical discrimination, and hidden groups that fear retaliation.

- Facebook should consider three key types of risks: (1) exposure to harm on our platforms, (2) capacity to mitigate issues on platform or as they move offline, and (3) real-world consequences that may result from content on our platforms.

- In order to identify groups at high risk of being "outed" on the platform, we must take into account their cross-cutting vulnerabilities that differ by location and a number of other factors.

Policy Relevance

Vulnerability, from the broader literature and an analysis of Facebook's specific situation, reveals exposure on our platforms amplifies various risks to users depending on location and the reasons they are at risk.

Watts, Michael & Bohle, Hans-Georg. (1993). The Space of Vulnerability: The Causal Structure of Hunger and Famine. Progress in Human Geography - PROG HUM GEOGR. 17. 43-67. 10.1177/030913259301700103.

Evaluation Framework for Outing-Risk Groups
External Outreach

We spoke with 31stakeholders, including human rights organizations, pro expression groups, and members of outing-risk groups.

Evaluation Framework for Outing-Risk Groups
External Outreach

Key Themes:

1. Broad consensus on protecting against a broader set of harms

2. Outing may come in the form of coded language or symbols

3. Intent and credibility of claims are irrelevant

4. Protect public figures, but make exceptions for those who should be held accountable (e.g., politicians)

5. Protect family members & allies, as well as locations

Evaluation Framework for Outing-Risk Groups
Guiding Questions

Part I: Evaluation Process

- How should we conduct assessments?

Part II: Proposed Policy

- Who/What should we protect?

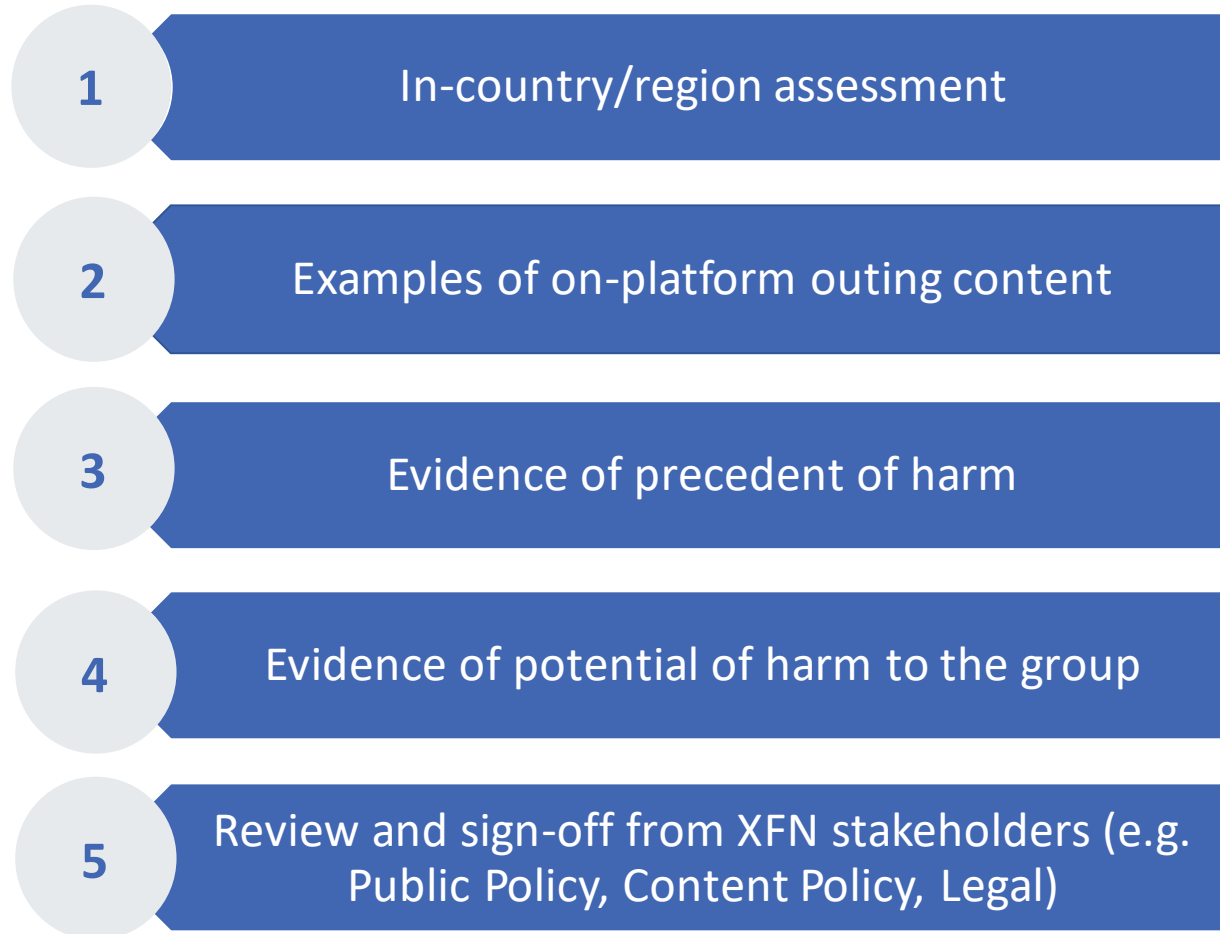- What types of outing should they be protected against?

Part I: Evaluation Process

How should we conduct assessments?

Evaluation Framework for Outing-Risk Groups
Evaluation Process

Content Policy conducts the designation process following the initial signal(s) that may come from a number of sources, including content review and public policy teams or external experts

1 In-country/region assessment

2 Examples of on-platform outing content

3 Evidence of precedent of harm

4 Evidence of potential of harm to the group

5 Review and sign-off from XFN stakeholders (e.g. Public Policy, Content Policy, Legal)

Designation process can be initiated by:
- Real world events (both current & historical)
- Input from internal stakeholders
- External feedback from partners
- Observed activity on Facebook and Instagram
- Media reports and external research

Conduct Annual Audits to confirm outing-risk status

Part II: Proposed Policy

Who/What should we protect and against what types of outing?

REMOVE: Content that exposes the **identity** of a **private individual** as a member of an outing-risk group
- *The current evaluation process is not standardized across various regions. As such, our current outing-risk groups list is inconsistent in the scope of protections offered.*

Pros

- Allows critical speech of public figures
- Easy to enforce at scale

Cons

- Only protects private individuals
- Limited scope of outing risks and harms
- Proof of intent may expose vulnerable people to harm
- Review based on intent is often subjective
- Inconsistent application

Evaluation Framework for Outing-Risk Groups
Examples – Option 1

Outing: Familial Relationship of a Political Figure (Apostasy in Sudan)	**Outing: Professional Activity of a Political Figure** (Blasphemy in Pakistan)	**Outing: Members** (LGBTQI in Egypt)	**Outing: Location** (Associated with an Outing-Risk Group)
Allow ✔	**Allow** ✔	**Remove** ✖	**Allow** ✔

"This is the brother of the Governor who renounced Islam. Look at him!"

"Minister Amjad is secretly advocating against capital punishment for blasphemers. Disgusted."

A page created with the sole purpose of exposing gays in Egypt

"Has anyone seen this Rehab Centre? They are protecting drug dealers. Duterte must see this!"

Part II: Proposed Policy – Option 2

Option 2

REMOVE: Content that exposes the **identity** or **locations** affiliated with **any individual** who is alleged to:
- Be a member of a group at high risk for outing; and/or
- Have performed professional activities in support of a group at high risk for outing

Pros

- Expands protection to address harmful outing content
- Protects outing of locations
- Easy to enforce at scale

Cons

- Removes speech concerning public figures, particularly politicians and government officials
- Does not fully address other types of outing (e.g., familial)

Evaluation Framework for Outing-Risk Groups
Examples – Option 2

Outing: Familial Relationship of a Political Figure
(Apostasy in Sudan)

Allow ✓

"This is the brother of the Governor who renounced Islam. Look at him!"

Outing: Professional Activity of a Political Figure
(Blasphemy in Pakistan)

Remove ✗

"Minister Amjad is secretly advocating against capital punishment for blasphemers. Disgusted."

Outing: Members
(LGBTQI in Egypt)

Remove ✗

A page created with the sole purpose of exposing gays in Egypt

Outing: Location
(Associated with an Outing-Risk Group)

Remove ✗

"Has anyone seen this Rehab Centre? They are protecting drug dealers. Duterte must see this!"

Evaluation Framework for Outing-Risk Groups
Part II: Proposed Policy – Option 3 (Recommendation)

REMOVE: Content that exposes the **identity** or **locations** affiliated with **any individual** who is alleged to:

- Be a member of a group at high risk for outing; and/or
- Share familial and/or romantic relationships with a member(s) of a group at high risk for outing; and/or
- Have performed professional activities in support of a group at high risk for outing **(except for political figures)**

Pros

- Expands protection to address harmful outing content
- Protects outing of locations
- Provides protection to politicians when they are members of an outing-risk group
- Allows speech about political figures related to professional activities

Cons

- Potentially restricts some speech related to public figures
- Difficult to determine local political figures

Evaluation Framework for Outing-Risk Groups
Examples – Option 3

Outing: Familial Relationship of a Political Figure (Apostasy in Sudan)	**Outing: Professional Activity of a Political Figure** (Blasphemy in Pakistan)	**Outing: Members** (LGBTQI in Egypt)	**Outing: Location** (Associated with an Outing-Risk Group)
Remove ❌	Allow ✅	Remove ❌	Remove ❌

"This is the brother of the Governor who renounced Islam. Look at him!"

"Minister Amjad is secretly advocating against capital punishment for blasphemers. Disgusted."

A page created with the sole purpose of exposing gays in Egypt

"Has anyone seen this Rehab Centre? They are protecting drug dealers. Duterte must see this!"

Evaluation Framework for Outing-Risk Groups
Part II: Proposed Policy – Options Summary

Options	Types of Outing	Scope of Protection
Option 1 (Status Quo)	• Member of an outing-risk group	• Private Individuals
Option 2	• Member of an outing-risk group; • Have performed professional activities in support of an outing-risk group	• All individuals • Locations
Option 3 (Recommendation)	• Member of an outing-risk group; • Have performed professional activities in support of an outing-risk group • Share familial and/or romantic relationships with a member(s) of an outing-risk group;	• All individuals for outing as a member • Locations • All individuals for familial and/or romantic relationships • All individuals **(except political figures)** for outing of professional activities and affiliations