POLICY FORUM

facebook

July 8, 2020

Agenda

- Recommendation: Obscene Sexualization of Public Figures
- Recommendation: Veiled Threats

Recommendation: Obscene Sexualization of Public Figures

Organic Content Policy

Issue

We aim to remove harassing content, particularly obscenely sexualized commentary that is intended to denigrate people. However, this type of speech may be used in political discourse and even in celebration of public figures. We want to explore whether we should provide more protection for public figures.

Obscene Sexualization of Public Figures

Overview

Recommendation

Remove some content that obscenely sexualizes adult public figures

Source
Feedback related to sexualized commentary against female public figures

External Outreach
70 External Engagements

Working Groups
6 XFN Working Groups

Obscene Sexualization of Public Figures

Status Quo - Adult Public Figures

✗ Protection for all individuals against:

- Repeatedly contacting someone in a manner that is sexually harassing
- Attacks based on a person's status as a victim of sexual assault, sexual exploitation, or domestic abuse
- Attacks through derogatory terms related to sexual activity (e.g., "whore", "slut")

✗ Protection for adult public figures when the content directly tags or is targeted at the individual:

- Statements of intent to engage in a sexual activity or advocating to engage in a sexual activity
- Claims about sexually transmitted diseases
- Derogatory terms related to female gendered cursing

✓ Allow content related to adult public figures that:

- Sexualizes another adult
- Makes claims about sexual activity
- Makes claims about romantic involvement, sexual orientation, or gender identity

Obscene Sexualization of Public Figures

Research Findings

- Unwanted sexualization is experienced by public figures of all professions as harmful/hate speech, harassment, abuse or bullying targeting their abilities, physical appearance, protected characteristics, or behavior
- Women and traditionally marginalized groups are targeted at a higher frequency and the unwanted sexualization can lead to high rates of physical, mental, and psychological disorders, especially for women
- Online anonymity encourages sexualization of this group and objectification can normalize worldwide violence against women and girls
- Some female and minority public figures see sexualization as agency over (and celebration of) their own body which has historically been regulated by external forces

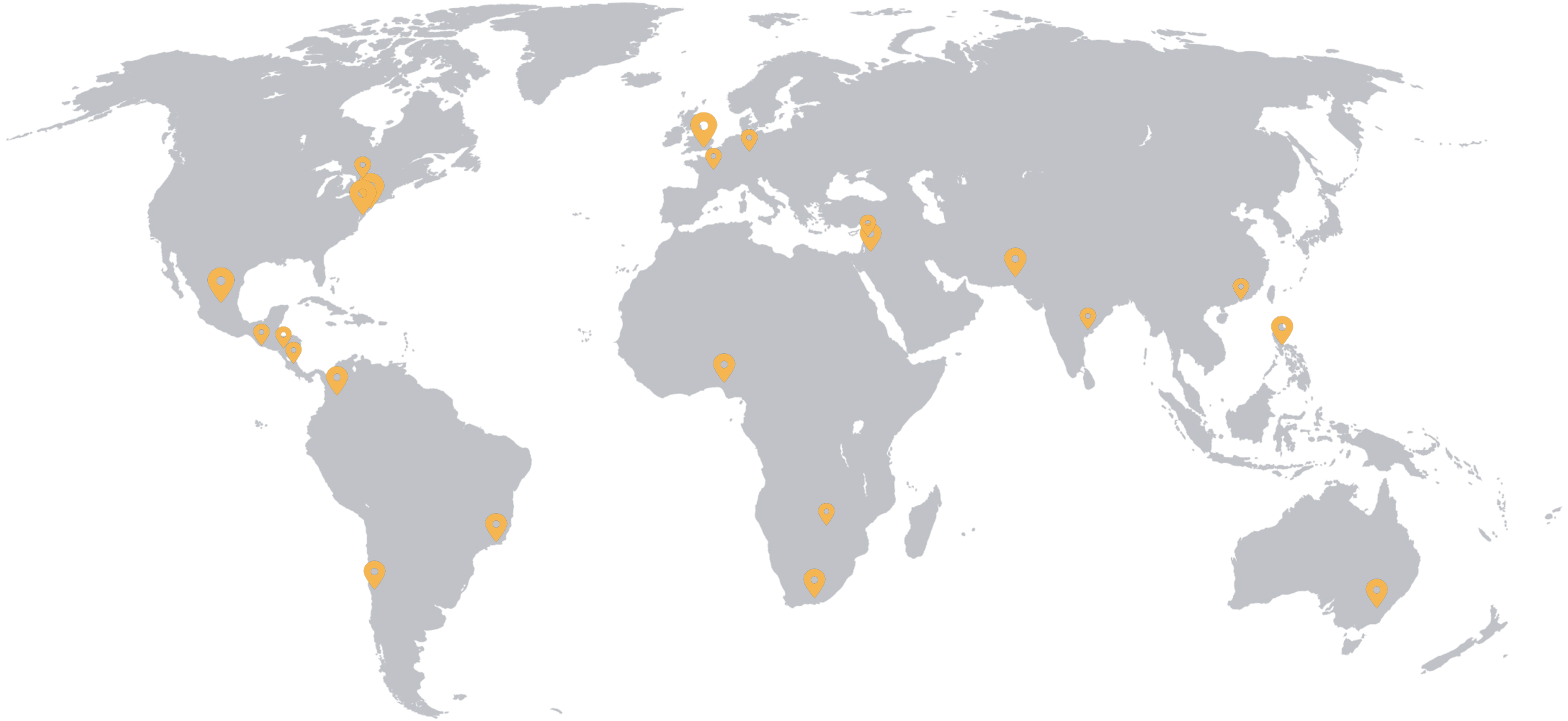
Sources: Internal research; Amnesty International (2020); UNICEF 2020; UNHR 2019; Kee (2016); Morris, Lynn, Goldenberg, et al. (2018); Karsay, Kathrin, Knoll, Matthes, et al. (2018); Council of Europe Network 2013; Sills, Sophie, et al. (2016); Rosenwald (2016);

Policy Relevance

Unwanted sexualization of public figures impacts users' health, voice, and safety which can drive negative online experiences and possible offline harm

Obscene Sexualization of Public Figures

External Outreach



We connected with 70 stakeholders on this issue, including academics, journalists, political organizations, women's rights groups, digital rights organizations and safety partners.

Obscene Sexualization of Public Figures

External Outreach

Key Themes:

- Sexualization disproportionately affects women, particularly those challenging historical power structures
- There is limited support for current broad definition of public figures

- Sexualization delegitimizes, silences, and in some cases endangers women
- Negative impacts reach beyond the target
- Such content curtails users' expression

- Repeated and coordinated harassment is seen as most problematic
- Sexualization is seen as more severe than commentary on physical appearance

- Unwantedness is an important factor but difficult to operationalize
- Pro-expression side and others recommend focus on user moderation

Obscene Sexualization of Public Figures

Option 1: Status Quo

Option 1

Remove attacks through derogatory terms related to sexual activity (e.g., “whore”, “slut”) &

Remove when the content directly tags or is targeted at the individual:

- Statements of intent to engage in a sexual activity or advocating to engage in a sexual activity
- Claims about sexually transmitted diseases
- Derogatory terms related to female gendered cursing

Pros:

- Allows most speech critical of public figures
- Operable

Cons:

- Does not remove potentially harmful speech
- This speech disproportionately affects the speech of women
- May have a chilling effect and lead to self-censorship

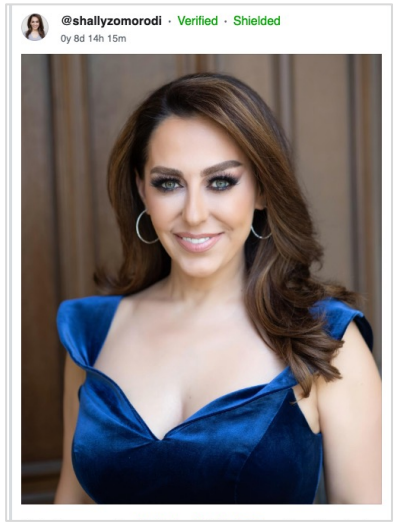
Obscene Sexualization of Public Figures

Examples: Status Quo

Sexual Orientation/ Activity Claims

... *"[Maria Ressa]had an affair with the producer, Lilibeth Frondoso, while she was in a long time relationship with another woman" and "**Beth fingered Maria in the office**."*

Describing Genitalia



"Perfect pussy 👌"
Or
"Huge dick"
Or
"Big balls"

Aspirational Statement

"I'd love to fuck you"

Statements of intent to engage in sexual activity

"I'm going to fuck you"

Visual Claim About Sexual Activity

Obscene Sexualization of Public Figures

Option 2: Narrow Scope

Option 2 (Rec)

Remove content that obscenely sexualizes individuals based on:

- Statements of intent to engage in a sexual activity with intended recipient of message, or advocating that intended recipient engage in a sexual activity

Pros:

- Removes explicit type of obscene sexualization
- Allows speech that is critical of public figures
- Operationally feasible

Cons:

- Does not remove other potentially harmful speech
- Allows speech that may create an environment of intimidation or self-censorship

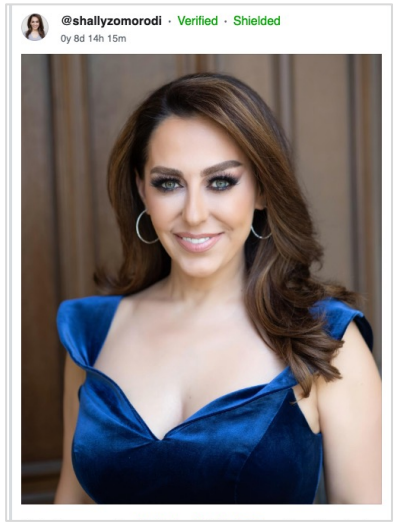
Obscene Sexualization of Public Figures

Examples: Option 2 - Narrow Scope

Sexual Orientation/ Activity Claims

... *"[Maria Ressa]had an affair with the producer, Lilibeth Frondoso, while she was in a long time relationship with another woman" and "**Beth fingered Maria in the office**."*

Describing Genitalia

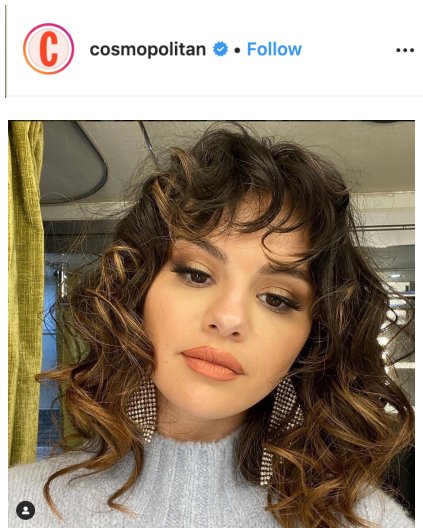


"Perfect pussy 👌"
Or
"Huge dick"
Or
"Big balls"

Aspirational Statement

"I'd love to fuck you"

Statements of intent to engage in sexual activity



"I'm going to fuck you"

Visual Claim About Sexual Activity

Obscene Sexualization of Public Figures

Option 3: Medium Scope

Option 3

Remove content that obscenely sexualizes individuals based on:

- Statements of intent to engage in a sexual activity with intended recipient of message, or advocating that intended recipient engage in a sexual activity
- Descriptions of genitalia (e.g., “big dick”, “perfect pussy”)

Pros:

- Addresses stakeholder and safety feedback
- Allows speech that is critical of public figures

Cons:

- Risk of over-enforcement and false positives
- Does not remove all potentially harmful speech

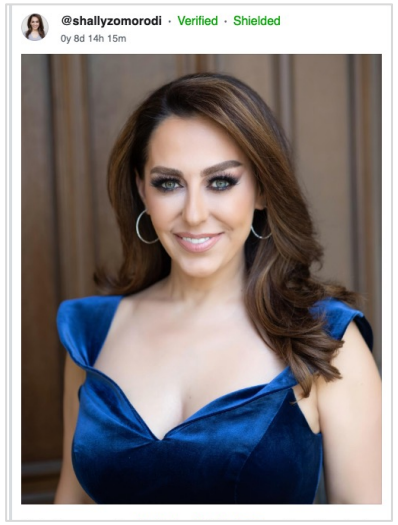
Obscene Sexualization of Public Figures

Examples: Option 3 - Medium Scope

Sexual Orientation/ Activity Claims

... *"[Maria Ressa]had an affair with the producer, Lilibeth Frondoso, while she was in a long time relationship with another woman" and "**Beth fingered Maria in the office**."*

Describing Genitalia

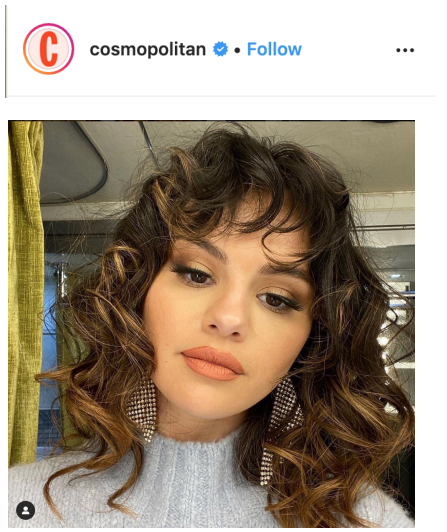


"Perfect pussy 👌"
Or
"Huge dick"
Or
"Big balls"

Aspirational Statement

"I'd love to fuck you"

Statements of intent to engage in sexual activity



"I'm going to fuck you"

Visual Claim About Sexual Activity

Obscene Sexualization of Public Figures

Option 4: Broad Scope

Option 4

Remove content that obscenely sexualizes individuals based on:

- Statements of intent to engage in a sexual activity with intended recipient of message, or advocating that intended recipient engage in a sexual activity
- Descriptions of genitalia (e.g., “big dick”, “perfect pussy”)
- Aspirational or conditional statements to engage in a sexual activity (e.g., “I want”, “I wish”)

Pros:

- Addresses stakeholder and safety feedback

Cons:

- Risk of over-enforcement and false positives
- May remove speech that is not unwanted
- Removes speech about public figures

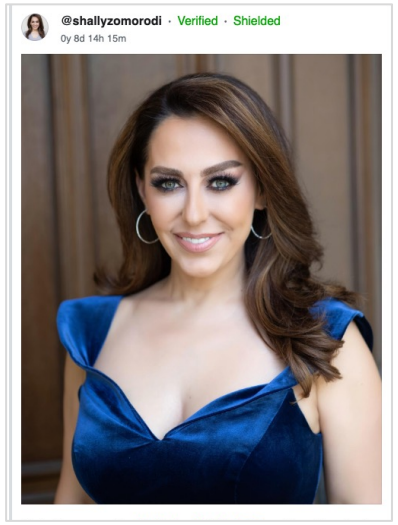
Obscene Sexualization of Public Figures

Examples: Option 4: Broad Scope

Sexual Orientation/ Activity Claims

... *"[Maria Ressa]had an affair with the producer, Lilibeth Frondoso, while she was in a long time relationship with another woman" and "**Beth fingered Maria in the office**."*

Describing Genitalia

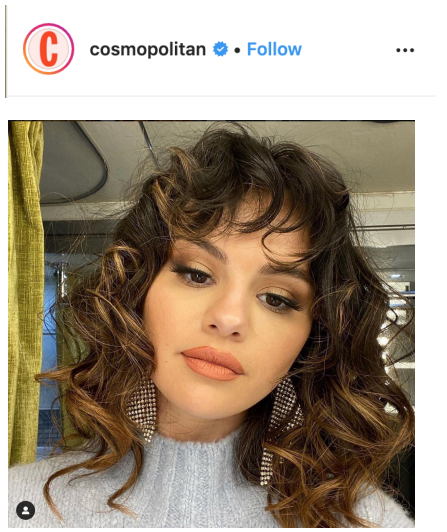


"Perfect pussy 👌"
Or
"Huge dick"
Or
"Big balls"

Aspirational Statement

"I'd love to fuck you"

Statements of intent to engage in sexual activity

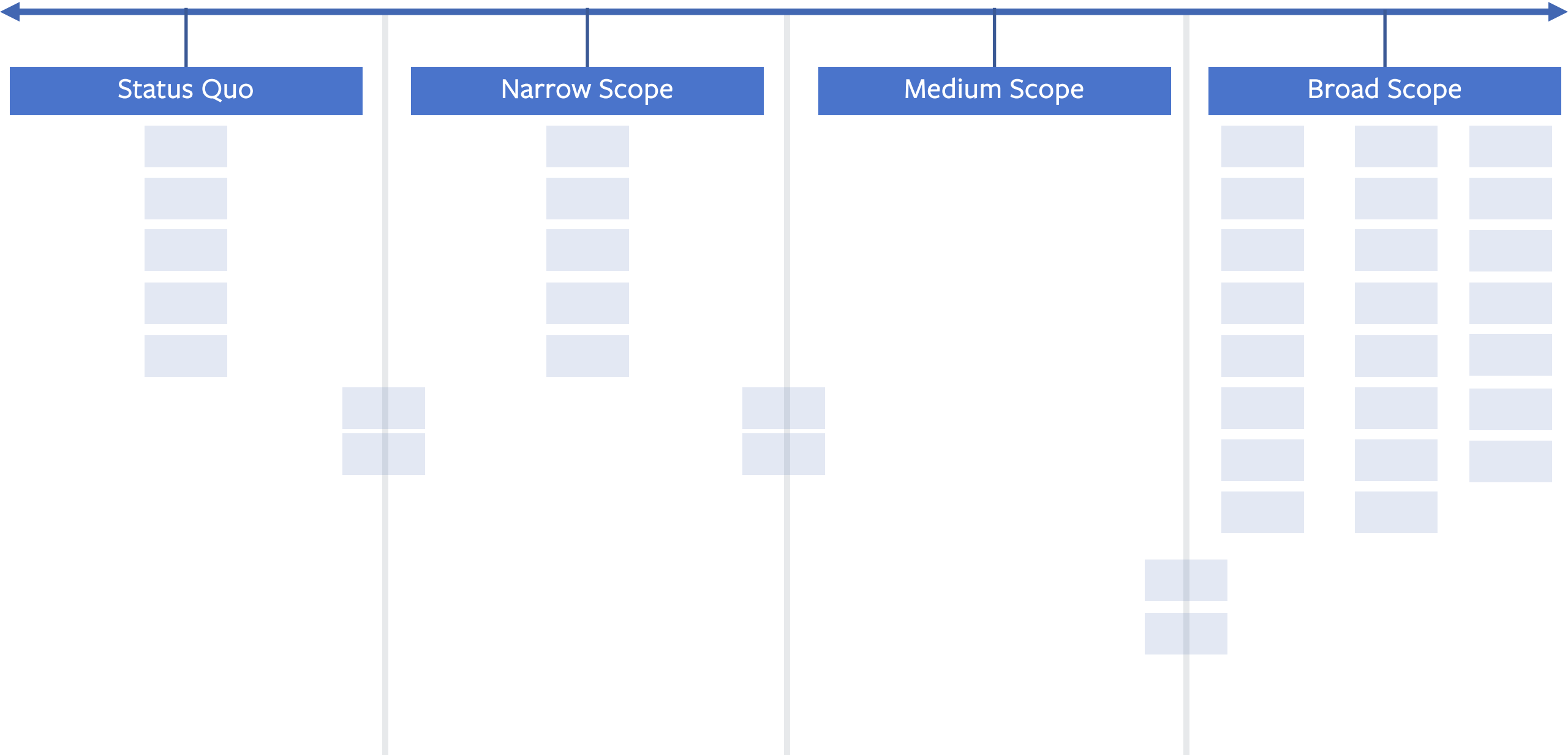


"I'm going to fuck you"

Visual Claim About Sexual Activity

Obscene Sexualization of Public Figures

External Outreach



Recommendation: Veiled Threats

Organic Content Policy

Issue

We remove explicitly violent threats under our Violence and Incitement policy, but we do not remove language people might *perceive* as threatening. We want to do more to remove veiled and implicit threats; however, the subjectivity of removing non-explicit threats would introduce inconsistency, bias and possible over-enforcement.

Veiled Threats

Overview

Recommendation

Improved assessment framework for escalations

Source

Inconsistent treatment of non-explicit threats and concerns from external stakeholders about perceived inaction on veiled threats

External Outreach

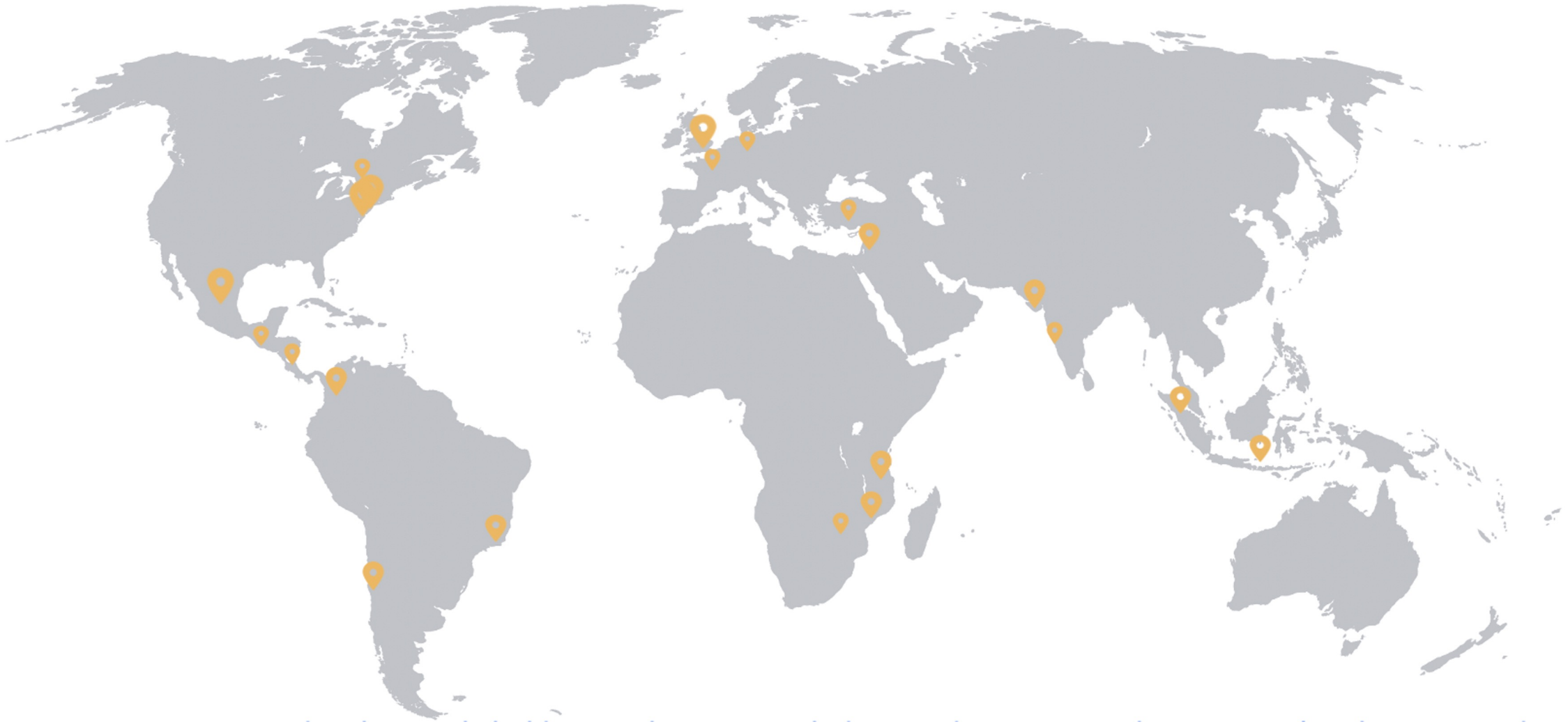
51 External Engagements

Working Groups

6 XFN Working Groups

Veiled Threats

External Outreach



We connected with 51 stakeholders on this issue, including academics, journalists, women's rights groups, digital rights organizations, and safety partners

Veiled Threats

External Outreach

Key Themes:

- Veiled threats' plain meaning do not express an intention to inflict harm, but an interpretation of the content, in its context, may express such an intention
- They are seen in various forms; their meaning depends heavily on the local context and regional trends that can change rapidly
- Veiled threats that are part of targeted misuse and organized campaigns are seen as most worrying by the stakeholders
- Stakeholders working on freedom of speech state that if the subject matter of the content is in any way within the public discourse, the causal link between speech and violence should be very clear

Veiled Threats

Status Quo

Option 1

No policy prohibition on threats that are not clearly articulated

Pros

- Clear parameters for enforcement already established
- Consistent in removing content that explicitly incites or facilitates violence

Cons

- Doesn't account for veiled threats with potential risks to personal and public safety
- In cases where we make spirit of the policy decisions, we may create the perception of inconsistent enforcement

Veiled Threats

Examples: Status Quo

Veiled Threat

Caption: "I'm in favor of the real female participants cornering this imposter afterwards and teaching him what it really means to be a woman"

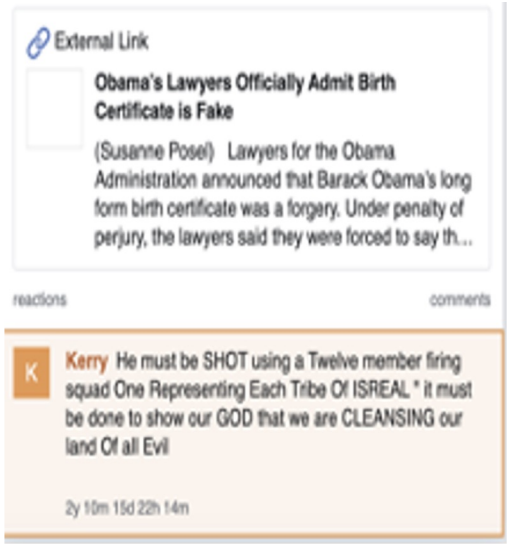
Veiled Threat

Text: "Server called me broke when I didn't tip her so now, I'm waiting by her vehicle until she gets off. We will see if she has the same energy"

Veiled Threat

'Remember this face' (plus an exact description of where his office is) and claims that he is a pro-Thaksin supporter

Explicit Threat



Comment underneath an article about Obama's alleged fake birth certificate: "He must be Shot using a twelve member firing squad"

Veiled Threats

Status quo + improved assessment framework for escalations

Option 2 (Rec)

On escalation, assess the presence and use of a veiled threat using established framework

Pros

- Holistic review allows for more effective enforcement of veiled threats that may create safety risk
- Establishes framework that can be used consistently on escalation

Cons

- Applying a framework on escalation may delay response time in cases involving imminent safety risk
- Limiting application of framework to escalations forestalls the possibility that we can address veiled through scaled review

Veiled Threats

Proposed indicators for identifying the presence of a veiled threat

Primary Indicators (At least 1 indicator required)	Secondary Indicators (At least 1 indicator required)
1. Content is shared in retaliatory context	A. Local context or subject matter expert confirms that the content in question is considered potentially threatening, or likely to contribute to imminent violence or physical harm.
2. Content references historical or fictional incidents of violence	
3. Content is acting as a threatening call to action	B. The target of the content reports the content to us, as verified by a name or face match report
4. Content shares sensitive information that could expose others to harm	

Veiled Threats

Create marketised lists of Veiled Threats

Option 3

- Create a list of designated Veiled Threats to capture dog whistles and proxy terms **solely** associated with violence in different countries/regions
- Build out a vetting and designation process for Organic Content Policy to assess items for inclusion on an ongoing basis

Pros

- Offers an at-scale enforcement option
- Factors in regional context without placing undue pressure on reviewers to identify Veiled Threats
- Lends itself to proactive detection of veiled threats

Cons

- Only covers a defined range of Veiled Threats
- Risk of over-enforcement
- Cognitive overload on reviewers as they have to adapt to utilizing another list

Veiled Threats

Employ a combination of Options 2 and 3

Option 4

- Status quo policy plus improved assessment framework for escalations
- In addition, launch a country/region-specific list of Veiled Threats for scaled review

Pros

- Two-pronged approach helps ensure increased impact
- Lends itself to proactive detection of veiled threats
- Multiple ways of adequately accounting for local context
- Factors in a clear assessment framework for making decisions on escalations

Cons

- Only covers a defined range of Veiled Threats at-scale
- Risk of over-enforcement
- Cognitive overload on reviewers as they have to adapt to utilizing another list

Veiled Threats

External Outreach


facebook