Roundtable discussion on Burma and Bangladesh with Meta

Thursday April 13, 2023 at 1130am

Re: Pilot Phase of the Framework for the U.S. Government's Engagement Related to Tech Platforms and Mass Atrocities/Other Key Events

---

Overall

- **Introduction and overview of the Initiative by USG**
- **Operating structure and Teams involved**
  - What is your current regional and in-country representation for Burma and Bangladesh?
  - What are weaknesses for your company's operations globally to address political violence or mass atrocities (e.g. language, detection, staffing, scale of issues, lack appropriate local contextual knowledge)?
  - What are the relevant teams and internal stakeholders involved in the prevention of violence around key events and atrocities? *(in line with table in Annex 1 to concept note re: Questions for Consideration in Engagement with Tech Platforms by Agencies, Bureaus)*
  - What types of third party vendors or actors are involved in the monitoring and enforcement of policies around prevention of violence and atrocities?
- **Policy**
  - ***From Burma team (USAID and State Pol/Econ):*** *"Atrocity" should be defined, if loosely, and potential perpetrators and victims identified. What best practices, industry standards and/or evidentiary tests are there that social media companies can use to define atrocity, perpetrator, and victim (especially in the context of complex inter-ethnic /inter-religious conflict like we see in Burma)?*
  - What policies or procedures do you have in place for addressing the use of your products/services for the commission or incitement of mass atrocities or political violence around key events?
    - What is the spectrum of policy responses your team considers? Are they responsive to factors like proportionality or likelihood, etc?
    - Where are these policies and procedures located? To what extent are they public or internal only?
    - Do these policies and procedures differ depending on your product/service?
    - What language are these policies and procedures provided in? *(Relevant to Burma and Bangladesh)*
  - What processes and policies do you have in place to ensure that online harms (abuse, harassment, calls for violence, and related disinformation) does not translate to offline harm?
    - How do you deal with regional or localized contextual nuance in global operations, and the issue of linguistic diversity and colloquial terms (e.g. hate speech terms about specific ethnic group in specific dialect.)
  - In enforcement of your policies re: key events and atrocity prevention, what is the time frame within which your team can adapt to new languages and/or regions? Or behaviors that are contributing to violence or atrocities?

Commented [1]: Would be great to touch base as State/USAID beforehand to revisit the outcomes and questions to be answered in the concept note to make sure we've got a robust approach to drive at those set of issues. I attempted to cross reference but think helpful if we discuss in advance.

Commented [2]: Information inflows to USAID should be aggregated at a level that minimizes any PII and avoids disclosure of any proprietary or trade secret information.

Commented [3]: USAID/GC: Reiterating that if purpose and expectations are not clearly articulated, these conversations may be misconstrued as indirect government pressure to take down content (regardless of content moderation policies/standards) and censor freedom of speech. This might become more challenging against the backdrop of certain democratic openings or events, such as elections. It might make sense to consider additional guardrails for ensuring that the framework does not have the unintended effect, in countries where such platforms are a primary means of engaging voters, of influencing or being perceived as influencing the outcome of the elections by taking down certain content. Per USAID's Political Party Assistance Policy, USAID does not seek to pick winners and losers and tries to provide equitable amounts of assistance. Incidents where protected free speech rather than hate speech was restricted as a result would pose reputational risks and could serve as fodder for PRC pushback on past USG criticism of PRC censorship or a narrative that the USG itself engages in media censorship.

Commented [4]: Will the introduction or other communications with Meta make it clear that participation in the roundtable discussions is voluntary and is not legally required?

Commented [5]: -USAID/GC: Please provide the talking points for the overview and introduction of the initiative.

-One practical concern is that platforms may be reluctant to candidly engage or treat this as the precursor to a US government investigation if the USG does not appropriately frame its questions.

- Keep in mind that 21 STATE 36050 places strict limits on engaging online platforms regarding content moderation so State and USAID should ensure these questions and its approach are consistent with that guidance. The cable underscores ...

Commented [6]: Have Comms been taken into consideration? If Meta announces this partnership publicly (Headline: "USG and Meta exchange user information to fight atrocity crimes") or decides to inform users that their information was disclosed to a USG agency, is the USG comfortable with this?

Commented [7]: Merged as related to org's operating structure

- What is your company's commitment to transparency regarding the enforcement of your policies re: atrocity prevention and key events (including processes, procedures internally)?
- How does your company communicate with users around policies and procedures, particularly for atrocity prevention and key events? Proactive or reactive?
- In addressing and preventing political violence and mass atrocities, how does your company factor in privacy concerns for individual users?
- How does your company address issues of potential bias or conflict of interest in responding to potential enforcement of policies and procedures dealing with mass atrocities or political violence?
- Benign content in one context may incite violence in another. Does your company have policies that allow for adaptiveness and responsiveness to crises or conflict situations (including political violence, evolving events that could be considered mass atrocities)?
  - Are there considered deviations from normal processes and procedures on such issues?
- What policies/procedures do you have in place to quickly pivot or surge resources to address the outbreak of violence in a region you don't already have a robust effort toward?

- **Specialized skills/backgrounds**
  - Does your company have individuals responsible for ensuring compliance with policies aimed at preventing the abuse of your platforms/services for the purpose of committing or inciting mass atrocities?
    - How about violence related to key events?
    - Where are those individuals embedded in your compliance structure?
    - What autonomy do those individuals have?
    - How are those individuals adequately resourced to act upon their mandate?
    - What authority do they have to escalate issues to senior leaders in the company?
  - Are there different approaches taken for different services offered?
- **Training**
  - What training do your employees receive, particularly as it relates to how your products/services can be used to commit or incite mass atrocities? What about relating to potential or actual offline harms related to key events?
    - Does it include engineering and product teams?
    - Do you provide similar training to contractors, temps, and vendors?
    - Is this training required or voluntary?
    - In what languages is the training provided?
    - What is the frequency of training and coverage re: geographic and cultural reach?
  - What type of communications approach does your company take in a lead up to key events where there is a potential for offline harms or atrocities (including engagements with civil society or academia)?
  - Are there specific trainings or engagements for relevant Staff related to Burma and Bangladesh?
- **Process to identify/monitor countries at risk**

- How do you detect, examine, and investigate acts aimed at abusing your platform/services for committing or inciting mass atrocities or violence around key events?
  - Do you report such misconduct to law enforcement or other relevant government authorities? Proactively in line with policies, or only as mandated under national law?
  - How do you detect such misconduct by foreign actors or internal threats?
  - How do you manage such misconduct by third parties?
- What processes do you have in place to identify, analyze, address, and share whether your products/services are being used for the commission or incitement of violence, offline harms or mass atrocities? What information or metrics does your company collect and use to help detect this type of misconduct? How frequently are these assessments revisited?
- What type of documentation of enforcement of policies and procedures does your company engage in for atrocity prevention and prevention of political violence around key events? Does this include escalations and external communications on specific instances?
- What incentives do you have encouraging employees, vendors, temps, and contractors for detecting and reporting abuses of your products/services aimed at inciting or committing political violence or mass atrocities?
- What disincentives do you have for persons who fail to detect and address such acts, such as factoring such abuses into performance evaluations or promotion opportunities?

- **Tech tools**
  - What types of interventions does your company use to prevent violence and atrocities? Does it cut across all types of services (eg, text, imagery, video, voice recordings/streamed engagements, groups)?
  - What type of continued analysis and risk mitigation (bias, inadvertent harm) goes into use of ML in enforcement of policies for violence and atrocity prevention?
  - Would it be possible to share and provide feedback on the list of classifiers (e.g. keywords and phrases) that are used to detect violence and atrocities at-scale, in local languages?
- **Local language contractors/subcontracting monitoring of platforms**
  - What is the capacity of Meta to enforce its policies in local languages for Burma and Bangladesh?
  - What type of training do language subcontractors receive on the context and policy issues around prevention of violence and atrocities?
  - Do you use third parties to detect and address the use of your products/services to commit or incite violence relating to key events or mass atrocities?
  - How do you ensure the third parties engaged in this work are properly qualified to undertake this work?
  - What systems do you have in place to ensure they are undertaking this work effectively?
- **Training**
  - What training do your employees receive, particularly as it relates to how your products/services can be used to commit or incite mass atrocities? What about relating to potential or actual offline harms related to key events?

Commented [8]: Needs more input from experts on this area

Commented [9]: Ideally, local partners who specialize in violence and atrocity prevention would be able to review and suggest edits to the list of classifiers, but if not even feedback from our local FSNs would be helpful.

Formatted: Font color: Blue

Formatted: Font color: Blue

        - Does it include engineering and product teams?
        - Do you provide similar training to contractors, temps, and vendors?
        - Is this training required or voluntary?
        - In what languages is the training provided?
        - What is the frequency of training and coverage re: geographic and cultural reach?
    - What type of communications approach does your company take in a lead up to key events where there is a potential for offline harms or atrocities (including engagements with civil society or academia)?
- **Information channels; Stakeholder Engagement with civil society (local, international)**
    - What form of information channels (e.g. news, human rights reports, trusted partners, escalation pathways) do your policy and enforcement teams consider as it pertains to atrocity prevention and key events?
    - How does Meta engage with civil society or other partners, local and international in both preparatory ways and dynamic, rapidly evolving situations? Specifically for Burma and Bangladesh? Specific to themes of atrocity prevention and violence? Can you share more about the trusted partners program?
    - What form of external stakeholder engagement does your company engage in to inform the development, implementation and enforcement of your policies and procedures, particularly as it pertains to offline harm around key events and atrocity prevention?
- **Documentation of enforcement/steps taken**
    - ***From Burma team (USAID and State Pol/Econ):*** what aggregated data are social media companies already sharing or could foreseeably share with the likes of State and USAID, and INGOs like Amnesty and Global Witness for atrocity prevention and mitigation?
    - What processes do you have for addressing exceptional circumstances with the propensity for harm requiring urgent prioritization re: violence prevention around key events and atrocity prevention?
    - Do you have a process for reporting and investigating circumstances where your products/services are used for committing or inciting mass atrocities or violence relating to key events?
        - How is this reporting mechanism publicized to your employees, temps, vendors, and contractors?
        - Do you have a process through which users can also submit complaints for similar issues?
        - Who are the teams that investigate these allegations?
        - Do your investigation teams have expertise in how technology can contribute to or amplify the commission of offline harms or mass atrocities?
- **After action process**
    - What is the systemic approach your company uses to learn from violence around key events or potential atrocities where the platform's services were utilized or played a role?
    - How are decisions from Meta's Oversight Board that touch upon violence or atrocity prevention impacting sustained policy and operational improvements for Meta? Eg: https://oversightboard.com/news/927673894608838-oversight-board-upholds-meta-s-original-decision-case-2021-014-fb-ua/
    - What process do you have to test controls aimed at preventing or mitigating abuses of your products/services aimed at committing or inciting mass atrocities or political

Commented [10]: Moved this up - as stakeholder engagement is very similar often to info channels

violence? Do you test those controls as part of your internal audit? How are the results of that audit actioned upon to ensure continuous improvement of your controls?
    - To what extent is external stakeholder engagement and external research related to your company's operations considered in continuous improvements around inciting mass atrocities or political violence?
- **Key Events -** Which events? How decided?
    -
- **Information gaps – Where can USG support?**
    - What type of information might the U.S. Government hypothetically provide that may better assist with your efforts to mitigate the risk of your platforms being used to incite violence?

Other (relevant internal questions and issues for USG - State and USAID - during Pilot)

- ***From Burma team (USAID and State Pol/Econ):*** Glad to see "operating structure" as the first agenda item. There is some concern that this initiative may be duplicating efforts or communication channels/ processes typically led by other offices at State. We've also already seen that communication between the State desk and post has been inconsistent/unclear. We would appreciate clear guidance on the recommendation reporting relationship/ feedback channels, considering the complexity emerging from the fact that we essentially have 4 units "at post" - State in Yangon, State at DSU, USAID in Yangon, and USAID at DSU.
    - *Note, operating structure in agenda refers to Annex with questions, and org structure of Meta, not USG organization, so important that we clarify this during the pilot*
- ***From Burma team (USAID and State Pol/Econ):*** *Given the origins of this initiative, we want to reiterate (as is reflected in recent reporting from post) that we don't anticipate elections in the near term, and certainly none that the USG would recognize, so we want to ensure framing doesn't center around elections.*
- ***From Burma team (USAID and State Pol/Econ):*** *Although it's still early days, do we have a sense of the anticipated or sought outcome(s) from this initiative?*
    - *In the concept note we articulated the following potential outcomes of engagement via the Framework for USAID/State:*
        - *Improved understanding of the methodology used by tech platforms to assess the risk of violence and mass atrocities and what tech platforms are and are not doing at a global level to prevent or mitigate atrocities or violence related to key events.*
        - *Improved prevention and response to challenging issues such as: online harassment and abuse (with a gender lens, noting highly gendered nature), disinformation and misinformation in select foreign markets, transparency, and accountability on global and country specific challenges, opportunities, and good practices.*
        - *Improved information sharing and awareness raising of potential risk for harm related to key events and atrocity prevention with tech platforms.*

Commented [11]: USAID/GC: Please flesh this out prior to the meetings with Meta to ensure sensitivities and guardrails are appropriately taken into account.

Commented [12]: Needs additional input for USG talking points about how we are going to talk about this with external stakeholders, particularly given classification questions and sensitivities involved. Important area for guardrails. *Flag for GC

Commented [13]: Do we want to open up for wider feedback from tech platforms about how productive engagements have been on these topics since it is happening at post/mission?

Commented [14]: USAID/GC: I recommend that the USG, of course, be in listening mode and not make any commitments regarding information sharing.

        - *Strengthened ongoing technical level engagement with tech platforms by USG stakeholders through framework level engagement.*
        - *Strengthened approaches to USG foreign policy and development programming on key events and atrocity prevention due to improved understanding and appreciation of tech platforms policies, systems, and processes, especially at a national and local level. Relatedly, will help to improve human rights due diligence and reputational risk management for USG engagements (programmatic and policy).*
        - *Message Administration prioritization of preventing violence related to key events and mass atrocities to private sector partners.*
    - *Would be good to refine and discuss in line with engagements with platforms.*
- ***From Burma team (USAID and State Pol/Econ):*** *Strong recommendation to name this initiative something clear and succinct to make coordination and communication more straightforward (we've been saying 'tech company engagement on atrocity prevention thing' which is a bit of a mouthful!).*
- ***From Burma team (USAID and State Pol/Econ):*** *As Post continues to regularly liaise with social media companies, how can the USG incentivize companies to modify their community standards and fair use policies when it is in a U.S. foreign policy interest, absent FCC regulations or federal statutes?*
- ***From Burma team (USAID and State Pol/Econ):*** *USAID/Burma is currently working on an assessment to respond to Congressional staffer engagements re: atrocity prevention in Burma (primarily focused on exploring the potential of support to early warning/early response interventions), and also a currently unrelated effort, at staffer invitation, to propose language for the upcoming FY 24 Section 7043(a) 'Burma specific' appropriation provision, which will provide funding to implement the authorizations of the Burma Act.*
- ***From Burma team (USAID and State Pol/Econ):*** *Post would appreciate clarity on potentially thorny constitutional questions that are going to be hard to answer without greater elaboration by Congress in the Communications Act, Communications Decency Act, or other Congressional statute that relates to what could be perceived as interference in free speech/expression platforms. Absent law enforcement requests predicated on federal statutes, how can social media companies balance privacy concerns with sharing user data with federal agencies, including State and USAID? What level of aggregation of user data (or how feasible to remove personal identifiers) to allow for trend analysis in the context of atrocity prevention and mitigation?*

Attachments:

- ***Tab 1:*** Draft Concept Note for Pilot Phase of the Framework for the U.S. Government's Engagement Related to Tech Platforms and Mass Atrocities/Other Key Events
- ***Tab 2:*** Annex 1: Questions for Consideration in Engagement with Tech Platforms by Agencies, Bureaus

Commented [15]: @cayogu@usaid.gov - To check with USAID/GC
Assigned to cayogu@usaid.gov

Commented [16]: Is this question internal to the USG or for Meta? If it is for Meta then USAID/GC has serious concerns about this question. What does "modify their community standards and fair use policies when it is in a U.S. foreign policy interest" mean? This portion of the question potentially raises reputational risk for the USG and could be used to fuel arguments that the USG is censoring or controlling media or provide fodder for comparison to PRC actions limiting or controlling free and independent media.

If it is for Meta, this question implies that the purpose(s) of this meeting is to pressure companies to change their community standards and fair use policies. More importantly, the last clause "absent FCC regulations or federal statutes" implies that the USG is seeking to do an end run around the typical regulatory processes.

On the basis of what statutory or regulatory authority would State or USAID be asking this question given that fair use and community standards do not typically fall under USAID or State's legal authorities?

For example, some might draw a negative inference from the fact that agencies that do not regulate this space are posing this question as opposed to the FCC conducting consultations on this issue or being included in these conversations.