

Content Standards Forum

February 12, 2019

TOPICS: Violations of International Humanitarian Law, Accusations of Violations of International Humanitarian Law

Agenda

- 1 DISCUSSION: Violations of International Humanitarian Law
- 2 HEADS UP: Accusations of Violations of International Humanitarian Law

Violations of International Humanitarian Law

International Humanitarian Law

Issue:

Currently, our policies prohibit content that praises or supports hate crimes, including genocide. In addition, any person convicted of genocide by an international court or tribunal qualifies for designation as a hate figure, and we remove all praise, support, and representation of hate figures. We want to consider whether we should further incorporate principles of international humanitarian law, such as war crimes and crimes against humanity, into our standards. However, this would put Facebook in the position of removing content praising, supporting, and representing state actors.

Summary to Date:

- Consulted with 22 external stakeholders.
- Convened 6 working group meetings.

Key Considerations

- Address violations on a global basis
- Establish an enforceable standard for crimes and actors
- Set a clear evidentiary threshold that accounts for developing situations
- Define enforcement options (will we remove violators only, or will we also remove praise, support & representation?)
- Determine whether to treat heads of differently than state actors given implications for political speech
- Develop framework to better understand how dangerous speech by state actors may be a precursor to IHL violations

To date, we have convened a series of internal and external working groups in order to determine how best to incorporate principles of international humanitarian law into our content policies. We hope to use the first part of today's Content Standards Forum to discuss, and get input on, the work we've done to date so that we can develop a recommendation and more clearly delineate next steps.

There is overwhelming consensus – internally and externally – that we should expand our policy to account for war crimes and crimes against humanity. There is less certainty on the specific evidentiary basis that would warrant such action.

Before we open it up for discussion, we also want to note that whatever we do here will be a first step, and that we will continue to iterate on our policies.

Discussion

An option that leans on the evidence put forth by international institutions (e.g. ICC, UN fact-finding missions) means that:

- We would be relying on advisory standards that are not universally recognized – and in some cases, are subject to criticism (e.g. The Goldstone Report);
- We are bound to a system that is slow and therefore is not responsive to

- developing situations around the world; and
- We risk inconsistency in the evidentiary standard we require since UN Fact-Finding Missions present their findings and recommendations based on a “reasonable belief” standard, whereas the ICC bases its decision on “clear and convincing evidence / beyond a reasonable doubt.

As an alternative, we could consider approaching this the way we do designations of dangerous organizations and individuals. Those are designations that we make on our own subject to a rigorous process that we have refined with input from external experts. If we go down this route, we avoid endorsement of international institutions and the evidence they put forth.

There were a few external experts that endorsed the idea of us coming up with our own internal process, criteria and standards to designate and remove representation of actors responsible for international crimes. But, for the most part, that was the minority view and most of the external stakeholders we spoke to recommended that we lean on external sources of information that are internationally recognized and respected (i.e. reports by UN-mandated Fact Finding Missions and Commissions of Inquiry as well as independent local and international NGOs).

As work on this policy proposal continues, not only must we work to address the current disparity between state and non-state actors, but we also need to account for the way different people and groups around the world are using our platform. In Myanmar, for example, we don't always see clear hate speech violations; instead, we see hateful speech propagated over time and by networks of bad actors.

As we said at the beginning of this discussion, the policy recommendation that we ultimately arrive on will be the first of a series of actions. The heads up that we're going to discuss next is something that has come up in our working group discussions – how should we treat accusations of violations of international humanitarian law.

Underlying all of this work is the work we are doing to better understand how dangerous speech by state actors may be seen as a precursor to on-the-ground crises and violence. We will plan to continue this discussion in internal and external working groups, and will reconvene with a specific policy recommendation once we've done more research and due diligence.

HEADS UP: Accusations of Violations of International Humanitarian Law

Accusations of Violations of International Humanitarian Law

Overview

Issue:

We are currently working on a policy that would likely prohibit individuals convicted of war crimes and crimes against humanity from having a presence on the platform. Relatedly, we want to think through how we might treat accusations of crimes against humanity or war crimes. We also want to explore how accusations of this sort could be used to trigger investigation into on-platform activity. In absence of a conviction, however, any action we take against accused actors must be based on strong principles and a rigorous process.

Goals:

- Convene a Working Group, including key stakeholders in Product Policy (Content, News Feed, Commerce) to consider different potential enforcement actions (e.g. demonetization, downranking) based on suspected violations of international humanitarian law.
- Work with our Cybersecurity and Threat Intel teams to determine when accusations may warrant an investigation into on-platform activity.

Accusations of Violations of International Humanitarian Law

Questions to Consider

- Barring removal, what are the different enforcement actions we can take against individuals suspected of committing war crimes or crimes against humanity?
- How should we handle content praising or supporting individuals suspected of committing war crimes or crimes against humanity?
- What evidence should be required in order to take action (e.g., UN mandated Fact Finding Mission, Special Rapporteur report, or a combination of local and international media)?
- Should this threshold also trigger a threat intelligence/coordinated inauthentic behavior investigation?

Accusations of Violations of International Humanitarian Law

Next Steps

- Schedule working groups
- Begin outreach to external experts

facebook