

Subject	Comments on Facebook Oversight Board Proposal
Date	June 4, 2019

Summary

The following document addresses Facebook's Oversight Board proposal around three elements: (i) The practical tension between international standards and Community Standards; (ii) the challenge of scale and context in content moderation, and (iii) the importance of the appeal process. Based on this points, we consider that the Oversight Board proposal –and its related debate– is overlooking key issues around the problem of content moderation in Facebook. In particular, we believe the Oversight Board should not be drafted as an 'international tribunal', but focused on advising and overseeing the implementation of the Community Standards. In that sense, we believe the proposed framework for the Oversight Board is inadequate.

1. The Oversight Board's mandate: Between international standards and Community Standards

According to Facebook's Draft Charter for the Oversight Board (OB), *"The board will be a body of independent experts who will review Facebook's most challenging content decisions - focusing on important and disputed cases"*. Furthermore, the Draft outlines three main responsibilities for the OB:

- Provide oversight of Facebook's content decisions.
- Reverse Facebook's decisions when necessary.
- Be an independent authority outside of Facebook.

According to several authorized voices within civil society, the OB's mandate should be guided by international standards and, specifically, recognized freedom of expression principles. Access Now, for example, made the following recommendation in its [comments about the OB proposal](#): *"It is necessary for all companies to apply these essential human rights principles when they make*

decisions that impact user speech, regardless of whether the decisions are made in-house by content moderators or externally by an oversight board”.

The concrete way these human rights principles would be implemented –and reflected– in both the OB mandate and the Community Standards (CS) is an overlooked issue by many civil society groups. Would the implementation of his framework affect, for example, the way Facebook balances users’ security and trust against freedom of expression? And, moreover, is Facebook willing to change an approach that currently restricts forms of speech that are legitimate under human rights instruments?

As Daphne Keller [explains](#), *“Even for committed free-expression advocates, it is not clear that requiring platforms to preserve all legal speech is in the public interest. There are speech rights on all sides of the issue”.* Also, there are clear economic incentives for platforms to reject a framework that abides by international standards: *“Requiring platforms to carry speech that most users don’t want to see would also have serious economic consequences. Among other things, platforms would lose revenue from advertisers who do not want their brands associated with hateful or offensive content”*, concludes Keller.

As the OB proposal stands today, Facebook has not explained how –or if– it expects the OB to reconcile the platform’s CS with international standards. By the same token, civil society groups are suggesting adjudication and due process guidelines for the OB, akin to international courts, disregarding the challenges of keeping online legitimate, but very problematic, speech. These disjointed visions fail to address the underpinning issues of content moderation and the best use of an OB within Facebook’s current reviewing process.

While this document does not attempt to suggest a structure or mandate for the OB, we believe that such a body should focus on assessing the existing enforcement of the CS, taking into account that this process is producing considerable false positives in the basic levels of moderation that are not being properly reviewed in further stages. In that context, the OB could also be an advisory instance on policy development and enforcement, but not an ‘adjudication’ body.

All in all, creating an instance –such as a ‘tribunal’– that is not articulated with the practical problems of content moderation at inferior levels, is inconvenient. In addition, it creates the risk of having an elite stage of content moderation that is useful only to a handful of users.

2. Content moderation: context and scale

According to Facebook, about thirty-thousand reviewers have the task of assessing content that can violate the Community Standards (CS) –either because it was directly reported by a user or detected by the platform’s tools. This endeavour unfolds in a service with 2.37 billion [Monthly Active Users](#) who post over a billion photos on a daily basis.

Facing the scale of this challenge, Facebook has underscored the need to have policies that are principled, explicable and operable. The latter aspect is paramount to Facebook’s moderation scheme: The enforcement of the CS have to be fast and scalable. But what about context? How can a massive moderation scheme incorporate nuance, local tensions and extra information in its review process?

“The need to balance between consistency of rules, with being sensitive to local contexts, particularly for issues like hate speech and disinformation, is of particular concern when considering both the design of platforms and issues of scale”, explains [Robyn Caplan](#). While small platforms can have processes that afford contextual assessments –either artisanal or community-based strategies–, Facebook, Twitter or Youtube, among others, can only introduce such kind of review in very specific cases.

We do not have any details related to the number of cases the OB would review. However, it seems clear that such a body will have a rational amount of cases, giving them precisely the time and space to implement a reviewing process that day-to-day moderators do not execute today and are unlikely to execute in the future. Apparently, the OB will deliberate, introducing information about local settings and case specifics. Without a doubt, this instance will be able to produce sound decisions around CS enforcement. However, what would be the use of such body if that process cannot be replicated at lower levels of content moderation? Besides some recommendations and concrete suggestions, the OB’s outcomes will only impact the concerned users.

3. The importance of the appeal process

The speed and scale requirements of Facebook’s industrial content moderation approach produces many false positives. This is not a problem the system can correct, but an outcome of its design. Hence, the reviewing process of such decisions is key to its legitimacy. On April 2018, Facebook announced an [appeal process](#) for its users. However, there is little information about its implementation,

and in Latin-American markets several users regularly voice their concerns around the uselessness and opacity of the process (not to mention the language localization problems).

The implementation of the OB risks to distract the need to develop and fortify the general appealing process. For instance –as this document suggests– the OB should focus on overseeing this issue. While first-level moderators cannot afford introducing contextual elements in their review, appeals can have more nuanced assessments under specific criteria. This stage should be transparent and accountable, and unfold under clear policies of mandatory review and avenues for users to have their content restored and accounts un-suspended. Following the DMCA model, the appeal process should correct ‘false positives’ produced in lower stages, but specifically it should allow the protection of the user’s voice in the platform.

* * *

About us

Linterna Verde is an interdisciplinary group devoted to exploring and understanding the digital public debate. Our goal is to build bridges to strengthen deliberation, promote critical citizenship, and illuminate the work of civil society. Linterna Verde is a non-profit organization registered in Colombia. This brief was written by [Carlos Cortés](#).