



WHAT THE FACEBOOK OVERSIGHT BOARD MEANS FOR HUMAN RIGHTS, AND WHERE WE GO FROM HERE

What the Facebook Oversight Board means for human rights, and where we go from here

JUNE 2020

Table of contents

Introduction	2
Where we are now	2
Potential for positive impact	3
Limited to a narrow selection of content	4
The role of law	4
The first batch of members and their regional impact: not all that glitters is gold	5
A vulnerable foundation: Facebook should consider a more rights-respecting business model	7
Conclusion	8

Introduction

In May 2020 Facebook [announced](#) the inaugural batch of members of its [Oversight Board](#), an initiative Mark Zuckerberg first hinted at in [April 2018](#) during an interview with Ezra Klein. The Board will be a group of 40 people — of which [20 have been recently announced](#) — supported by staff, whose initial task will be to serve as an independent appeals mechanism to have a final say on select cases of content removals decided by Facebook. This is an important development that will have significant implications for how Facebook — and [others who may follow suit](#) — make decisions about what can be shared on its platform.

Access Now has been tracking this initiative closely and has provided analysis, commentary, and recommendations — including a [full report](#) on content governance and the proposed Board in 2019, and our [comprehensive guide](#) on how to protect human rights in content governance in 2020. We have also participated in multiple in-person stakeholder discussions and roundtables on the matter.

Below we take a high-level look at this first iteration of the Board, its potential for impact, and challenges that lie ahead. Much more can be said about its functional challenges, and we defer to the detailed analysis that other experts such as [Evelyn Douek](#) have performed on that front. We also place the Board in context, and highlight the importance of navigating its relationship with regulatory frameworks and Facebook’s overall business model. Finally, we identify the immediate next steps Facebook should take to safeguard freedom of expression and other human rights.

Where we are now

When a piece of content on Facebook — a photo, post, or comment, for example — is reported by a user or detected through [automated screening](#), it is reviewed by the company’s moderation system using both automated tools and human decision-making. Once Facebook’s moderators decide to remove a piece of content, the user that uploaded or created it — the content producer — can appeal, referring the case to another human moderator for a final decision. Usually, here is where the process would end. But once the Oversight Board begins operating, users will have another opportunity to escalate content removal decisions for external review. The Board, through a case selection committee, will choose novel cases to review that can provide guidance on the interpretation of Facebook’s community standards.

It took a year and a half for Facebook to create this entity, including a [public consultation](#) on the structure and functioning of the Board, targeted stakeholder [engagement](#), and a [selection process](#) for the first batch of members. Access Now participated in the stakeholder consultations by joining in-person discussions and through a [written submission](#).

With the creation of the Oversight Board, Facebook intends to add an instance of independent appeal for some content takedown decisions the company makes based on its terms of service. By doing so, it also aims to bring some form of legitimacy into content moderation decisions that affect billions of users worldwide.

However, achieving that legitimacy is going to be a major challenge for Facebook, and depends on the Board's ability to respond with deep nuance and cultural sensitivity — something Facebook has often struggled with thus far — and its ability to adapt where governments develop new regulatory frameworks — though Facebook has demonstrated a clear preference for self-regulation.

And while the Board can provide improvements for some of Facebook's most pressing content governance issues by addressing high-profile or precedent-setting cases, it needs to be allied with better on-the-ground responses to daily content matters as well as a broader human rights review of Facebook's business model. The Board also needs to iterate and to be prepared to fail. It should recognize that this is a real-time experiment with very real human consequences.

Potential for positive impact

Through its products Facebook, Whatsapp, and Instagram, [Facebook Inc.](#) dominates the market of social media platforms. Facebook, at its sole discretion, makes the rules about what kind of content to allow on its platforms through its community guidelines, how it distributes that content, and how those rules should be interpreted and applied, affecting the safety and rights of billions of people worldwide. And complex trade-offs come into play when a dominant private business that provides a platform for expression tries to make consistent [decisions at scale](#).

Facebook's Oversight Board constitutes a novel approach to content moderation that has the potential to bring [much-needed](#) external independent assessment to cases that can have a significant [impact](#) on freedom of expression. According to the Board's foundational documents — the [charter](#) and the [bylaws](#) — the Board will work independently and will publish the rationale and details of its decisions which, in turn, will be mandatory for Facebook. In addition, Facebook will be obligated to respond to the Board's advice publicly. The charter also establishes that the Board's decisions will pay special attention to the impact of international human rights norms protecting free expression.

To ensure its independence, the foundational documents establish that the Board will set a case selection committee with absolute discretion on what cases to hear. It will operate under an [independent trust](#) that has been funded by Facebook to ensure the operations of the Board and its dedicated staff for 6 years.

In this respect, the Board could be an innovative non-judicial grievance mechanism to provide some form of remedy to affected users, in line with the recommendations contained in the [United Nations Guiding Principles on Business and Human Rights](#) and in our [Remedy Plan](#).

Limited to a narrow selection of content

Despite this potential, the Board’s decision-making power — and therefore its ability to make meaningful and sustaining impact — is limited. In its first stage, it will only be able to review decisions to take specific content down, putting aside decisions to keep content up and everything related to groups, pages, events, and fact-checking until some unspecified time in the future. The Board also will not have a say about the workings of recommendation algorithms or decisions to demote or promote content or ads — which is a tool used to deal with [misinformation](#), for example. Arguably, most of Facebook’s impact on public discourse derives from the company’s opaque algorithmic amplification practices, both in the platform newsfeed and with regard to ads.

The Board will focus only on the free expression dimension of content removals, but Facebook’s human rights impact goes way beyond that. Facebook’s data harvesting business model — built on opaque algorithmic recommendation for engagement — has serious impacts on free expression as well as privacy and community safety, and may build negative incentives for public debate. As such the Oversight Board is a welcome development, but it needs to be seen in a broader context and set within a larger set of questions that ought to be addressed in order to resolve the human rights challenges that continue to plague Facebook.

The role of law

An important question for the near future will be how Facebook’s Oversight Board will interact with national and regional law already in place and on the horizon. So far, Facebook has entrenched its dominance and fended off regulatory initiatives by self-regulating to the extreme of even proposing its own form of “[judiciary](#).” But we should be cautious of equating the initiative of a private platform with the robust democratic processes that are necessary to protect human rights, and we cannot let this approach distract us from the real problem: that a private actor has unprecedented power over public discourse.

When it comes to laws that are already in place, the [bylaws](#) state that the Board will not decide on removals based on legal obligations or a “valid report of illegality.” The Board also will not decide on content that is unlawful in a jurisdiction connected to the content — i.e. domicile of the producer or the reporting user.

This regime sounds right in terms of leaving the Board out of the enforcement of laws and policies instituted by democratically elected governments. That should be the role of judicial systems.

Facebook, as a member of the [Global Network Initiative](#), already has [committed to performing](#) a careful legal analysis and performing human rights due diligence when it comes to government requests for content removal. This analysis allows Facebook to avoid complying with orders that are based on laws or authorities that are at odds with basic human rights — an objective that is difficult to uphold in practice alongside [business interests and draconian regulations and enforcement](#).

But what happens when regulation goes one step forward? Just like Facebook’s own Community Guidelines and Terms of Service do, legislation and different regulatory initiatives could go beyond illegal content into the realm of “harmful but legal” content. What happens when new regulations establish procedures for increased accountability, oversight, and procedural obligations for content removals? Several countries and regions are [preparing legislation](#) that, for better or worse, may deal with these issues in different forms — e.g. [duties of care](#), oversight mechanisms, and co-regulatory measures — that may collide with the work of the Board.

Facebook promises that the Board will follow the law by default. But there is a real risk that future procedural obligations based on legislation are sidestepped by evaluating removal requests on the basis of terms of service or [“community guidelines” first](#). In fact, Germany’s Federal Office of Justice says that situation happened with the procedural obligations under the [controversial](#) German “Network Enforcement Act” (NetzDG). Facebook’s underreporting of the number of complaints it had received about illegal content on its platform resulted in a 2.3 million euro [fine for Facebook](#) in July 2019.

The Board intends to put [international human rights law and freedom of expression first, but it’s also bound to make its decisions by interpreting Facebook’s privately-set content policies and values](#). If self-regulation keeps advancing without paying attention to legal developments, Facebook’s terms of service and community guidelines could become the *de facto* procedural rules to determine content takedowns — even when there is regulation in place. And the Board may become just a mechanism of strengthening and reinforcing Facebook policies.

Moreover, several of those involved speak of the Oversight Board as a project that begins with Facebook but may extend to other social media platforms, which could similarly outsource content-related decisions to this entity. If that comes to pass, the Board may gain outside influence in human rights decision-making over content, further excluding other stakeholders.

The first batch of members and their regional impact: not all that glitters is gold

The newly minted Board is [supposed to be diverse and representative](#), bringing together experts who are “from different professional, cultural and religious backgrounds and have various political viewpoints.” At first glance, the first announced batch tick a number of those boxes. When you zoom in, however, the picture becomes less assuring and more worrying.

The Oversight Board is just one piece of the entire system that goes into determining how information is shared on Facebook's platform, spanning teams dedicated to content review to those focused on product development, policy, and more. Successfully responding to the unique needs of the many communities who engage on Facebook requires ensuring that voices from our communities are a meaningful part of every stage of the process.

For decades, the tech sector has heard calls to diversify its workforce, including in the 2001 Durban Declaration and Programme of Action of the World Conference against Racism, Racial Discrimination, Xenophobia and Related Intolerance. Yet in 2019, [only 3.8 percent of Facebook employees](#) in the U.S. identified as African American.

While Facebook tried to remedy this problem with its Oversight Board, the first 20 members do not represent the user base. As others have [pointed out](#), five of the board members are from the United States while 90% of the users are outside the U.S., and there are no representatives of other non-geographical communities such as LGBTQ people, indigenous people, and people with disabilities, among others.

Further, individuals should be selected for their expertise on content moderation issues, and not as regional tokens. Some members of the board are well-trusted in our communities, and have a stellar record of expertise on digital content issues and international human rights frameworks. But that is not the case for all the individuals selected, and members from the Middle East in particular set off alarm bells.

The selection of Yemeni activist and Nobel Prize winner Tawwakol Karman was met with strong backlash on social media mainly because of her political affiliations. While part of the backlash is an [orchestrated online smear campaign](#), many activists and civil society organizations in the region still have serious concerns about Karman's ability to unequivocally uphold the human rights of users given her lack of expertise on digital content governance and freedom of expression.

Palestinian human rights organizations, on the other hand, condemned [the selection of Emi Palmor](#), the former General Director of the Israeli Ministry of Justice. During her time in the role between 2014 and 2019, the ministry established the Israeli Cyber Unit, whose task is to [submit private requests](#) to social media companies for content removal. Consequently, the unit has been responsible for actively removing tens of thousands of Palestinian users' content with the number of requests jumping from 2,241 in 2016 to 14,283 in 2018.

For a region that sparked much of the early discussions on freedom of expression and social media following the Arab Spring in 2011, these two appointments — in combination — are bad news. Not only because they demonstrate a lack of understanding of the region's nuanced content moderation challenges, but also because they highlight flaws in the selection process itself where public

consultations stop at the vetting process, and the vetting criteria for weeding out conflict of interest seem to be insufficient.

While on the whole many of the Board's first members are excellent, well-respected choices, this was a missed opportunity to represent communities at risk, particularly in the MENA region where censorship, suppression, and criminalization of freedom of expression have escalated to unprecedented levels.

A vulnerable foundation: Facebook should consider a more rights-respecting business model

Facebook gathers and analyzes information about how we engage with its platform — and often, our activity beyond the platform as well — to build detailed profiles about our behavior, interests, beliefs, and more. The more information we generate, the more accurate the profiling, which is a very valuable resource for advertisers who pay Facebook to place highly targeted ads. Researchers — and even [Facebook itself](#) — have observed that content that elicits strong [emotional responses](#) generates more engagement. This is true for [political ads](#) as well. This is due to [psychological biases](#) common to all humans, but an algorithm designed to select content to increase engagement is also likely to promote content that [fosters outrage or prioritizes falsities](#).

The Oversight Board may eventually review cases centering around this type of content. But beyond investing resources into responding to the content itself, Facebook should also direct those resources toward reviewing the structural model that has helped such content thrive on the platform.

It's worth highlighting that, regardless of the merits of the Board, Facebook should not delay or defer structural changes long-needed at the company. We have [called for Facebook](#) to “Reform the Board of Directors, including for more diversity, expertise in data and human rights, and oversight of risks to users.” The company remains accountable to its CEO, Board Chair, and top shareholder, who has little human rights expertise or experience, and the staff of the company does not represent its broad, global user base. We hope this Board does not distract from greater efforts to [diversify Facebook](#).

But the incentives of a massive ad-driven company will rarely be plurality, diversity, and meaningful information exchange, and we cannot count on Facebook or its competitors alone. Therefore, democratic debate and rights-respecting regulatory action — including data protection, free expression safeguards, user safety, and competition, among other areas — can help provide more power to our communities and improve content oversight.

Conclusion

Overall, the issue of content governance in a major dominant platform involves hard questions and complex trade-offs. Facebook has a responsibility to uphold human rights and prevent harms by providing its users with transparency, notice, review mechanisms, and access to remedy — across all its products and activities. The Oversight Board can help provide that when it comes to content takedown decisions. But Facebook’s impact on freedom of expression and other fundamental rights is significant, persistent, and goes well beyond what it has tasked the Board to deal with.

To that end, if the Board proves to be effective and successful in its endeavor, Facebook should grant a wider scope to review additional categories of content, such as ads or fact-checking, and other product features, like algorithmic recommendation or demotion. This would allow the Board to take a **more comprehensive approach** to its oversight role, looking beyond just the publishing of user content and investigating the relationships between content, product functionality, and advertising streams.

Further, it is the responsibility of the Board co-chairs — together with Facebook — to [select new members](#). They should **do better in the next round of Board appointments** and more proactively consult civil society groups to improve evaluations of candidates' commitment to human rights, their specific expertise, and the overall diversity of the Board.

While the impact of content moderation on freedom of expression is obvious, both content decisions in concrete cases and the platform’s operations more broadly have a huge impact on other internationally recognized human rights, including on the right to privacy. Facebook should **establish and expand human rights safeguards and expertise** at the level of the corporate board, executive leadership, and throughout the company — including through a [global independent audit](#) of its data practices.

Addressing the data exploitation model and Facebook’s market power will be crucial to building a healthier digital public sphere. The [company’s power — and that of other Silicon Valley giants — will likely increase even during the COVID-19 crisis](#) due to the [suspension of probes](#), the leverage of dominance to launch [new services](#) such as Facebook Shop, and even new acquisitions like Facebook’s [buyout of Giphy](#).

A **wider, richer, and more complex debate** is needed to provide solutions that are compatible with human rights and that bring democratic governance and accountability into play. We expect governments to live up to the task, generating the space for evidence-based, participatory, and rights-respecting frameworks for content governance. We need to restore the significance of powerful democratic concepts such as “legitimacy,” “constitution,” and “supreme court” so we don’t use them as [metaphors](#) for company practices, but rather in reference to the essential building blocks of

democracy, which are the tools we need for reaching a stronger, safer, more rights-respecting online space for everyone.

This paper was authored by Javier Pallero, with contributions by Marwa Fatafta, Carolyn Tackett, and members of the Access Now Policy Team.

For more information, contact:

Javier Pallero (javier@accessnow.org)

Please direct media inquiries to

press@accessnow.org



Access Now (<https://www.accessnow.org>) defends and extends the digital rights of users at risk around the world. By combining direct technical support, comprehensive policy engagement, global advocacy, grassroots grantmaking, legal interventions, and convenings such as RightsCon, we fight for human rights in the digital age.