CONTENT MODERATION AND SELF-REGULATION MECHANISMS

The Facebook Oversight Board and Its Implications for Latin America

Edison Lanza and Matías Jackson
The Inter-American Dialogue’s Peter D. Bell Rule of Law Program is pleased to present the report Content Moderation and Self-Regulation Mechanisms: the Facebook Oversight Board and its implications for Latin America. This new report is an important contribution to the ongoing work of the Dialogue on freedom of expression and content moderation.

Just as the Internet has continued to evolve, so too have the methods for the regulation of online speech. The need for content moderation on social media platforms that respects democratic norms and abides by international human rights standards has led to a debate among policymakers, civil society, platforms and academic experts on the best approach towards this issue.

The Facebook Oversight Board, established in 2020, is an innovative approach for reviewing the content moderation decisions of Facebook and Instagram. An independent body that has already published its first group of decisions, the Board is an experiment in self-regulation by one of the most powerful social media platforms. The impact that such a body could have on the region merits close attention.

This report analyzes the following: (1) how the Board operates, (2) the first set of decisions the Board has published, (3) how the Board applies international human rights standards, (4) the role of the judiciary on content moderation, and (5) the impact on the regulatory discussion in Latin America. The conclusions and recommendations outlined here are an initial attempt at addressing some concerns voiced by regional civil society groups and how to better incorporate international human rights standards into the Board’s operation and decision-making.

This report was authored by Edison Lanza, senior fellow for the Peter D. Bell Rule of Law Program and former Special Rapporteur for Freedom of Expression at the Inter-American Commission on Human Rights, and Matías Jackson, a Uruguayan lawyer and Assistant Professor on Technology Law for the University of the Republic of Uruguay.

The Dialogue is grateful to the authors of the report for their important work on this critical topic, and to Facebook for its support of the Peter D. Bell Rule of Law Program.

MICHAEL SHIFTER
President, Inter-American Dialogue

SANTIAGO CANTON
Director, Peter D. Bell Rule of Law Program, Inter-American Dialogue
# Table of Contents

Executive Summary

I. Introduction

II. A Sui Generis Response for Growing Challenges
   First Set of Decisions

III. Application of International Law and Standards of the Inter-American System
   Considerations on the Facebook Oversight Board

IV. The Role of the Judicial Branch in the Internet Era

V. Impact on the Regulatory Discussion in Latin America

Conclusions and Recommendations

Governments

Facebook Oversight Board

Platforms

Civil Society

Endnotes
This report describes the main points of interest regarding the operation of the Facebook Oversight Board (FOB) and the inclusion of international human rights law in its decisions, with special emphasis on its impact on freedom of expression in Latin America. It closely examines some new aspects that this appeal mechanism presents in response to complex or controversial decisions of content moderation carried out by the main platforms for the circulation of information.

The competence and qualifications of the FOB’s first members, as well as its first set of decisions, have generated expectations about the Board’s potential scope of action. Academic and civil society organizations that defend digital rights have weighed in on the incorporation of international human rights standards and instruments, and the attention to local contexts in the Board’s decisions.

Nevertheless, the mechanism has its limits and cannot be expected to tackle or resolve all of the complex problems associated with the circulation of information in the digital era. Moving forward, doubts persist about some of the unilateral powers the company maintains in the mechanism’s founding charter, as well as the potential impact on the heart of the very business model that might be the cause of problems affecting freedom of expression and democratic systems in the digital era.

This analysis also covers the viewpoints and challenges observed by experts and civil society regarding the nature of this type of self-regulation mechanism; it takes an in-depth look at both the impact that the FOB and other non-state mechanisms could have on local judicial decisions and at the legal discussions taking place in the region around content regulation on digital platforms. The report also addresses the need for dialogue between the region and the Inter-American Human Rights System.

This mechanism should not be expected to be the only form of accountability in response to the growing impact of these private companies on public debate. Due to its administrative costs and complexity of operation, this mechanism is not transferable to just any internet platform. For this reason, stakeholders in internet governance, civil society and industry associated with these platforms should explore other appeals mechanisms adjusted to different contexts.

In parallel, several countries in the region have been considering legislation on content moderation. Several pieces of legislation were analyzed for this report, from which we can conclude that even when pursuing legitimate ends, such legislation is in general not aligned with global and regional standards on freedom of expression, as reflected in questioning from the international community and local organizations.

Finally, we provide recommendations for different stakeholders in the Latin American region to encourage public debate framed by international human rights standards (see Figure 1).
# FIGURE 1: SUMMARY OF RECOMMENDATIONS

## Governments
- Respect human rights standards
- Make decisions based on due legal process
- Promote multisectoral settings
- Analyze FOB decisions and other self-regulation mechanisms

## Facebook Oversight Board
- Incorporate regional standards
- Implement clear, contextualized consultation mechanisms
- Expand the range of jurisdictions served
- Expand complaint and interaction mechanisms
- Establish accountability mechanisms

## Platforms
- Undertake long-term commitments to transparency
- Comply with international standards
- Explore new mechanisms
- Commit to the decisions and recommendations
- Engage in dialogue with other companies

## Civil Society
- Continue the exchange of ideas with the FOB
- Foster greater participation
- Monitor government initiatives
I. Introduction

The internet is an open, decentralized medium when it comes to who may upload, download and share information. This has led to the development of digital platforms specialized in mediating access to information and online debate.

The model of openness and free circulation of information and opinions has shaped the internet since its commercial development in the mid-1990s. The limited liability of intermediaries, as well as the neutrality of the network with regard to data origin, content and flow have been core pillars for the development of innovation and pluralism on the internet.

Nevertheless, the technological reality and expansion of complex phenomena, such as the increasing spread of disinformation, hate speech driven by discrimination, or surveillance by means of digital tools, have added complexity to the application of standards forged in previous decades to protect the right to free speech.

More than twenty years after the expansion of this medium, we are facing a process of discussion or revision of these paradigms. There is a growing debate at different levels about the regulations applicable to these spaces, with positions that range from promoting self-moderation by the platforms, to proposals to establish models of state regulation. Meanwhile, non-government actors and human rights organizations are calling for transparency and commitment from internet companies to apply the standards of protection of international human rights law in their decisions.¹

Largely as a consequence of these challenges, Facebook put in place an independent review mechanism for its content moderation decisions. The Facebook Oversight Board (FOB) was established as an independent review mechanism and financed by the company with a trust of $130 million USD. In 2020 the company appointed four co-chairs, who in turn designated the remaining 20 board members based on a set of previously established criteria.²

Since January 2021, the FOB began to deliberate and publish a first set of decisions that may have significant consequences not only on debates about the role of internet platforms, but in shaping the content moderation policies and standards of free speech online. This is especially true given that Facebook is the world’s largest digital media platform, which operates in all countries and connects nearly 3 billion people.³

Furthermore, this is not the only self-regulation mechanism in place. In June 2020, Twitter announced that it will publish a series of principles to make its own content moderation more transparent, while at the same time adopting stronger decisions for regulating the spread of messages inciting hate speech or that could undermine democratic institutions. These decisions have been at the center of the debate, given that they have impacted world leaders and public figures and were adopted without jurisdictional oversight.

Google, in turn, has declared its adherence to both the Universal Declaration of Human Rights, as well as the Guiding Principles on Business and Human Rights. It is also part of the Global Network Initiative, a multi-stakeholder network to provide solutions in freedom of expression and content moderation policies; the company also periodically publishes transparency reports on content that has been removed or suppressed.

Google exercises automated content moderation on its platform by applying machine learning to detect content that is contrary to community rules or illegal according to each jurisdiction; however, it does offer the parties affected by these decisions the possibility to appeal and provide further information. The company
also established measures to foster the flow of quality information on its networks, especially YouTube, through the removal and reduced scope of potentially harmful or false information and, in addition, greater visibility and compensation for the creators of quality information. It has also developed diverse initiatives to promote independent journalism with investments reaching $189 million dollars in 118 countries.

All these initiatives are relevant to Latin America, a region where platforms have increasing influence on public debate and access to information. Facebook, for instance, has over 60 percent penetration among social media users in a region where many social, political and electoral events are mediated by discussion on these sites. Likewise, the problems stemming from automated content moderation by internet companies, deliberate spread of disinformation and polarization that in some cases may lead to violence, have also reached unexpected levels in Latin American countries.

In this setting, the need to create shared practices and principles in the region is undeniable when it comes to new forms of digital communication, bearing in mind that their scope goes way beyond the Latin American region or any single company.

The preparation of this report was preceded by a panel convened by the Inter-American Dialogue’s Rule of Law Program, the University of California’s International Justice Clinic at the Irvine School of Law, and Columbia Global Freedom of Expression at Columbia University. In May 2021, these three centers organized the event, “The Decisions of Facebook’s Oversight Board: Implications for the Global South, particularly in Latin America” with participation by international experts and Board members Catalina Botero and Jamal Greene. The following week, the Inter-American Dialogue also organized a high-level meeting with the OAS Special Rapporteur on Freedom of Expression and representatives from civil society organizations associated with the defense of online human rights to hear their views and comments on the functioning of the Board.

The need to forge shared practices and principles in the region is undeniable when it comes to new forms of digital communication, bearing in mind that their scope goes way beyond the Latin American region or any single company.
II. A Sui Generis Response for Growing Challenges

The creation of the FOB was preceded by a series of problematic incidents related to different content that circulate on social media – use of personal user data to influence elections, massive injections of deliberate disinformation and rhetoric inciting violence motivated by discrimination or even intended to generate chaos in the context of democratic election processes, among other disruptive phenomena.9

Faced with increasingly frequent incidents, fueled on social media by diverse actors and interests, many states realized they had to intervene. The US Congress summoned the CEOs of the major platforms to respond to these incidents. In the European Union, some states decided to regulate by law the circulation of content on these platforms that is limited or prohibited by the legal tradition of those countries, such as hate speech and discrimination.

In recent years, different civil society actors and international human rights organizations have recommended that internet companies put in place independent multi-stakeholder advisory boards to review their decisions as a form of accountability and to provide guarantees against private infringement on freedom of expression.10 Civil society, various networks of organizations and forums have also presented initiatives for smart regulation of different aspects of content moderation according to human rights standards.11

During 2019 and 2020, Facebook initiated a series of dialogues and exchanges to put the Facebook Oversight Board (FOB) into operation.12 The initial drafts presented by the company received objections from civil society at the global level, including the need to specifically incorporate human rights standards and ensure adequate levels of transparency and independence for moderation tasks.13 For his part, David Kaye, the United Nations Rapporteur on Freedom of Expression, sent a public letter to Mark Zuckerberg suggesting changes to align Facebook’s proposal with the UN Guiding Principles for Business and Human Rights.14

According to the FOB’s Charter, its purpose is to protect freedom of expression by making independent decisions and issuing recommendations on Facebook’s content policies.15 The Charter also sets forth the Board’s composition, scope, procedure and governance.16 The Charter is complemented by its Bylaws.17

BOX 1: SEEKING FINANCIAL AND OPERATIONAL INDEPENDENCE

From a financial standpoint, Facebook assigned a total of $130 million dollars for the first six months of the Board’s operation, covering operating costs and salaries for members and permanent staff. The funds are administered by the Oversight Board Trust, created specifically for that purpose, which “will guarantee that there is governance and accountability, and ensure that the Board complies with its stipulated purpose.”8

The Charter also determines the composition of the FOB, establishing the process of selection and removal. The initial selection of four co-chairs was done jointly with Facebook; the co-chairs then in turn selected renowned experts from around the world based on their expertise and track record in the defense of free speech and human rights. The FOB is currently made up of 20 members, including two representatives from the Latin America and the Caribbean region: Catalina Botero of Colombia and Ronaldo Lemos of Brazil. The members will serve for a period of three years, renewable up to a maximum of nine. In the future, selection of new members will be handled by a special committee of the Board, with an expected expansion of the FOB to a maximum of 40 members.
According to the Bylaws, Facebook or any of Facebook's or Instagram's users may submit a case for review by the FOB. Currently, the FOB admits two types of requests for review: 1) claims for reinstatement of posts that have been removed from the platform, and 2) claims to remove content that users consider must be eliminated. In order to resort to the FOB, users must first exhaust Facebook's internal review mechanisms.

Thus, a user who has appealed a decision that affects them will receive notification of Facebook's final decision and, if the content qualifies, they will also receive a reference identification number for purposes of review by the FOB. If that person is not satisfied with the outcome of their appeal, they may choose to refer their case to the FOB within 15 days of Facebook's final decision. Another path provided in the Bylaws is for the company to send the claim directly to the FOB.

An FOB Case Selection Committee of five members has the power to choose “which requests it will review and what decisions it will make in regard.” The Board's selection criteria are based on the difficulty and the significance of the case. Difficulty here refers to whether “the content raises questions about current policies or their enforcement, with strong arguments on both sides for either removing or leaving up the content under review.” The factors for consideration in terms of Difficulty include the possibility of applying various policies to the case, uncertainty as to whether the company's decision was made in accordance with Facebook policy, and the tension between equally important values. In the case of Significance, this means that “the content involves real-world impact.” The factors involved in Significance are “issues that are severe (threatening someone else’s voice, safety, privacy, or dignity), large-scale and important for public discourse.”

Once a case has been selected, deliberation will be handled by a panel of five Board members, including at least one representative from the region involved. Then a brief description of the case is published, opening a period of public commentary for 14 days. In addition, the members of the FOB may request information from Facebook, experts and outside organizations.

The decision adopted by the panel is then submitted to the rest of the FOB. If a special majority of the Board deems it necessary, the case could be referred to a new panel for an expedited review. Otherwise, the final decision is reported to the users involved and published on the FOB website. The decision may include policy recommendations for the company, which, although they are not binding, Facebook should respond to. The FOB will publish an annual transparency report including an analysis of how its decisions have taken international human rights law into account in the cases reviewed.

FIRST SET OF DECISIONS

In January 2021 the FOB published its first five decisions. The cases selected dealt with hate speech, dangerous persons and organizations, and erroneous information about health, sexual content, and discrimination. In four of these five decisions, the FOB recommended that Facebook review its initial decision to take the content down and ordered that the posts be restored.

Furthermore, each of these decisions included recommendations for the company to reform aspects of its content moderation policy. In this first set of resolutions, the FOB urged Facebook to provide more clarity about its Community Standards – including its relationship across different Facebook products – to develop more transparency standards in the use of artificial intelligence for content moderation and to better consider the contexts in which the material was posted. In late February, Facebook responded to these recommendations by committing to assess their viability and take action on most of them.
In a questionnaire sent by the authors of the report to the policymakers of Facebook’s public policies for Latin America, they explained that, once the company decides to “accept a recommendation from the Board,” it carries out “the policy changes and relevant compliance to implement it in the best way possible.” “We will publish an in-depth report on the commitments we have made as a result of the Board’s recommendations and provide a regional breakdown of the cases we have sent to the Board for its consideration,” they added.

The first set of decisions generated responses from academics, the media and civil society. The detailed inclusion of international human rights standards in these decisions was a salient aspect in general, as were the recommendations to improve the company’s content policies. Nevertheless, some criticized that the Board had selected cases involving “simple” issues and that the adverse decisions would not affect aspects of Facebook’s business model, such as the selection of information by algorithm or issues related to storage and use of users’ personal data.24

This first set of rulings includes a decision by Facebook that affects users in Brazil and concerns the elimination of a photograph showing a woman’s nipples.25 The image published on Instagram was part of an awareness-raising campaign about breast cancer symptoms in coordination with the international “Pink October” campaign. Facebook’s automated systems eliminated the photo for supposed infringement of its community norms on nudity and adult sexual activity. The FOB annull ed Facebook’s decision, concluding that the community norms do permit users to share nude photos when done “to raise awareness about a cause or [for] educational or medical reasons.” It also recommended that the company improve its appeal systems for automated detection and its transparency when it comes to reporting on their use.

In June 2021, the FOB announced the selection of a second case from the Latin American region, which again involved Brazil.26 It concerns a publication of the State Medical Council of Brazil on the effectiveness of lockdowns as a measure to reduce the spread of Covid-19. The content questions lockdowns as a measure for containing the pandemic, asserting that the measure has been condemned by the World Health Organization and causes a series of secondary consequences, such as alcohol and drug abuse. Facebook kept the post up because it did not violate its policy of erroneous or harmful information, but decided to refer it to the Board as a “complicated” case that could be used “for the adoption of certain safety measures during the pandemic.”

In July 2021 a third case from the region was announced, this time in Colombia.27 This is a case in which a media outlet shared a video that included text expressing admiration for those who appeared in it. The video showed a protest where people chanted insults against the Colombian president during demonstrations over tax reform in the country. Facebook decided to take the content down for violating its community norms on hate speech. The user appealed the decision, which the company chose to uphold; for this reason, the FOB mechanism was enabled.

In addition to the impact on the specific case, the interaction between the FOB and Facebook appears to open another interesting window into understanding how content moderation by a private entity operates. In the course of reviewing these cases, the FOB makes requests for information from the company that generate responses or evidence of a lack of transparency in some criteria, all of which allows for a better understanding of the moderation made by the platform and the most salient points of discussion.

In this first set of resolutions, the FOB urged Facebook to provide more clarity about its Community Standards - including its relationship across different Facebook products - to develop more transparency for content moderation and to better consider the contexts in which the material was posted.
BOX 2: THE DECISION IN THE TRUMP CASE

On January 6, 2021, during the Electoral College count to confirm the results of the 2020 elections in the United States, a mob assaulted the US Capitol in Washington, DC. Over the course of those events, then-President Donald Trump made two posts directed at the crowd that was violently entering the Capitol. In response, Facebook blocked access to his accounts for a period of 24 hours.

After reviewing the messages and collecting information on the events that occurred at the Capitol, Facebook decided to extend the block on his accounts “indefinitely.” Twitter, a platform on which Trump had made similar statements, made the same decision. For the first time ever, major social media networks indefinitely suspended the highest authority of a country. These unprecedented decisions generated debate around the world about the power wielded by these internet platforms over freedom of expression, public debate, and their influence on the societies in which they operate. Questions were raised as to the legitimacy of this type of decision, the responsibility of companies during events involving violence, and the application of international human rights law by private actors.

It was in this context that Facebook decided to send the case for study by the FOB. During the period of public consultation, the FOB received a total of 9,666 comments, 23 of which came from the LAC region. Finally, on May 5, in its announcement of its decision, the FOB ultimately agreed with Facebook’s position to restrict the access of Donald Trump to his accounts during the insurrection, considering that under the application of the Rabat Plan of Action, the messages of the then-president incited the protesters to continue their violent action. Nevertheless, it considered the indefinite suspension imposed by the platform to be inappropriate since it applied “an indeterminate and standardless penalty.” For this reason, it granted Facebook a period of six months to analyze the matter and provide a proportionate response.

In its response to the decision, Facebook committed to fully implement the FOB’s recommendations. This commitment includes the review of policies and response times to content from “highly influential” users that may cause imminent harm or pose threats to the integrity of people. The company also made the decision to uphold the suspension of Trump’s account for two years, after which the situation will be reassessed considering the existing risks at the time of reassessment.

Twitter, in contrast, decided to permanently suspend Trump’s account. Recently, in light of his policy of evading prohibition, the platform has shut down other accounts that claim they are not administered by the ex-president, but that exclusively share the content posted on his website.
III. Application of International Law and Standards of the Inter-American System

The need to guarantee the application of international standards for the protection of freedom of expression in content moderation is one of the main demands of different stakeholders who ask for limits on platforms’ discretion to determine the circulation, to invisibilize or to suppress speech or information.

In the words of Lee Bollinger, President and Seth Low Professor of Columbia University, there is no doubt as to the professional credentials of FOB members. However, the legitimacy of its members and their experience in the defense of human rights should not be confused with the Board’s institutional role, so it is worth asking what role the mechanism will play “in the creation of standards and norms on freedom of expression around the world.”

In this sense, Catalina Botero, a co-chair of the FOB, is optimistic about “the application of international human rights legal standards – universal and regional – in the moderation of online content” as the mechanism begins generating legitimacy and decisions. She added that this is not an automatic process, given the different legal traditions interacting within the Board and the global nature of the platforms under the FOB’s jurisdiction.

Jamal Green, also a co-chair of the FOB and professor of Constitutional Law at Columbia University, highlighted that there are certain dimensions that pose differences in the application of international human rights law on content moderation, depending on the different legal traditions. An example given by Greene is the different role that private citizens and the State play in the discourse that is disseminated through the public forum doctrine traditionally linked to the jurisprudence of the United States.

It is interesting to observe that the FOB itself, from its inception, has pursued the objective of legitimizing its mandate and the decisions it adopts in international law. Indeed, the Board’s Charter specifically includes that “the Board shall pay particular attention to the impact of removing content in light of human rights norms protecting free expression,” in addition to Facebook’s community norms that, for obvious reasons, should be considered.

So far, FOB decisions that have applied international law have preferred to incorporate standards from instruments (treaties, declarations and reports) from organizations in the universal system.

In practice, the Board’s decisions have incorporated standards established in international law from the beginning. Indeed, when the issues addressed are of high impact, such as discourse by high-profile public officials (see Box 2: The Decision in the Trump Case), the Board’s decisions have clearly been anchored by criteria, norms and definitions from the international legal context.

In the case of restriction of access to former-President Trump’s account, the Board indicated that “heads of state and other high government officials can have a greater power to cause harm than other people. If a head of state or high government official has repeatedly posted messages that pose a risk of harm under international human rights norms, Facebook should suspend the account for a period sufficient to protect against imminent harm.”

The Board also applied third-party review as prescribed by international law to evaluate Facebook’s community norms that restrict speech inciting violence on its platform, and applied the criteria of the Rabat Plan of Action to analyze whether the language in question had exceeded the threshold of protection afforded by freedom of expression to political discourse. Although it appeared to validate the immediate suspension of former President Trump’s account between January 6 and 7, 2021, the decision warns that the indefinite suspension of the account was inappropriate (see Box 2).

The decision underscored that Facebook had announced its corporate policy on human rights weeks before, in accordance with the UN Guiding Principles on Business and...
Human Rights. This criterion lays out the duty of private businesses to respect, guarantee and, as the case may be, redress fundamental rights violations, and that consist of “what businesses should do on a voluntary basis to meet these responsibilities” in terms of human rights, as well as to respond and prevent such violations when the business identifies harm done by itself or third parties.

It also draws from international treaties, such as the Covenant on Civil and Political Rights, recommendations by inter-governmental oversight organizations like the United Nations Human Rights Council, and even soft law norms, such as the Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence.

In a more recent case, related to debates on the platform on protests in Russia over the situation of leader Alexei Navalny, the FOB decided that, while the elimination of a series of comments was based on the community norm regarding bullying and harassment, it was an unnecessary and disproportionate restriction on the freedom of expression under international human rights standards.

On case in Brazil, Facebook’s automated systems eliminated a photo showing female nipples on the grounds that it violated the community norm on nudity and adult sexual activity. The FOB annulled Facebook’s decision to take down the post; in its view, the community norms allow users to share nudes when this is done to “raise awareness about a cause or educational or medical reasons.”

In making this distinction, the FOB cited norms from the International Covenant on Civil and Political Rights, the Convention on the Rights of the Child, and various reports by the UN Special Rapporteur on Freedom of Opinion and Expression. The Board also carried out a detailed analysis of the connection between freedom of expression and women’s right to health based on international law.

Although the FOB takes into consideration the context of the situation of vulnerability of women in much of Brazil, as well as the possibility that automated content moderation policies generate discriminatory impacts between men and women, the decision does not include references to norms, reports or decisions of the inter-American human rights system related to freedom of expression and women’s right to health, even though the region has robustly developed these standards.

In summary, so far FOB decisions that have applied international law have preferred to incorporate standards from instruments (treaties, declarations and reports) from organizations in the universal system.

The members of the FOB who participated in the initiative hosted by the Inter-American Dialogue agreed that the Board’s objective is to offer a universal lens on the right to freedom of expression, with the understanding that the universality and core principals of this right are of global aspiration. They also added that the socio-political context and legal traditions of each region are taken into account by the FOB through consultations with civil society organizations and experts in digital rights with broad expertise in their respective regions.

During the meeting organized by the Inter-American Dialogue with civil society organizations, it became clear there was a need to pay attention and open a space of dialogue with the FOB in order to have an impact on any discussion of content moderation. How standards on freedom of expression have been developed in the Latin American region, with its own perspective and traditions, must also be reflected in the FOB’s decisions.
American region, with its own perspective and traditions, must also be reflected in the FOB’s decisions.

CONSIDERATIONS ON THE FOB

The novelty of the review system established by Facebook, along with the high-profile figures involved in some of its decisions, has generated attention from governments, academia and civil society around the world.

In November 2018, Facebook CEO Mark Zuckerberg announced the creation of the FOB with the following comments: “First, it will prevent the concentration of too much decision-making within our teams. Second, it will create accountability and oversight. Third, it will provide assurance that these decisions are made in the best interests of our community and not for commercial reasons.”

However, the FOB is not free from critical analysis. As it is a non-state mechanism linked to the protection of freedom of expression, deficits have been noted in regard to its legitimacy, the range of available remedies, and its accessibility given the number of cases potentially affected by the decisions the platform adopts.

The most critical positions see the Board as an attempt to create a private Court of Justice to resolve matters related to fundamental rights that should be reserved for a state or international tribunal with democratic legitimacy. There is no lack of media platforms and experts who have pointed out the irony that Facebook established its own “Supreme Court.” A New York Times report asserts that the original idea came from Harvard law professor Noah Feldman who, during a series of meetings held with Facebook’s directors, argued that the companies that operate media or social networks should create “quasi judicial” systems to debate complex issues related to freedom of expression.

Other opinions situate the FOB in the tradition of ethics boards or media and journalism associations, which are ultimately self-regulation mechanisms on press content, recognized and encouraged by international law. In fact, several decisions by the European Court of Human Rights incorporate or refer to professional and ethical decisions of Press Associations. In the region, the IACHR Declaration of Principles on Freedom of Expression sets forth that, when dealing with content in communications media, it prefers rules on ethics or conduct over those imposed by the State.

As for the cast of remedies and reparations the FOB may provide, the approach adopted so far has been incremental: it began by accepting only cases over content that has remained online, with new areas expected to be added in the future.

Possible interpretations aside, the members of the FOB who participated in the Inter-American Dialogue event stated that they do not aspire to replace nor to create a supra-state judicial structure. Moreover, there is consensus among different actors on the competence and independence of the Board members, and that this type of mechanism can help to resolve some of the problems stemming from online content regulation.

Nevertheless, it has its limits and should not be expected to take on the entire set of problems associated with the circulation of information in the digital era. Experts like former Rapporteurs David Kay and Frank La Rue, pointed to the difficulty of replicating the FOB model among the rest of the internet industry. Only giants like Facebook have the financial backing to bear the costs of such a mechanism and, on the other hand, the pluralism of solutions would not help to generate replicable standards either. Thus, Kaye considers this idea, but warns that the FOB “must not monopolize” the entire discussion around content moderation; at the same time, he called for encouraging other panels and forms of evaluating the decisions made by internet-based companies.

Another line of reflection has to do with the level of independence the FOB will manage to build from the company that gave it its origin. There is still skepticism as to the unilateral powers the company maintains to modify some criteria in the Founding Charter and the fact it reserved for itself the initial task of selecting the first
four members. Some groups, such as the “Real Facebook Oversight Board” criticize the FOB for the selection of its members, done in conjunction with Facebook, and the lack of representation of geographic or identity minorities.

As for the cast of remedies and reparations the FOB may provide, the approach adopted so far has been incremental: it began by accepting only cases of removed content, then expanded to claims over content that has remained online, with new areas expected to be added in the future. The Office of the Special Rapporteur on Freedom of Expression at the United Nations has stated that this approach, although valid for this experimental stage, simply limits the cases that can be submitted for resolution, leaving [other] victims of human rights violations without remedy. Specifically, the Rapporteur points to cases that have not been addressed, such as political advertisements and the reduction of visibility of content by algorithms or geo-blocking measures.

From the standpoint of accessibility, the Bylaws have been criticized for not providing for the possibility that people who are not network users, but who may be affected by its content, can submit cases for consideration. According to the Bylaws, in order to request a case review by the FOB, the person must have an active Facebook or Instagram account and have exhausted all internal review mechanisms.

Some authors assert that the decision in the Trump case shows that the FOB is more like a human rights court or an international monitoring institution than a supreme court. Any new mechanism constructs its legitimacy gradually through a body of decisions, moving from general questions open to interpretation, to more concrete – and possibly more transformative – issues.

According to the United Nations Guiding Principles on Business and Human Rights, in order to ensure the effectiveness of non-judicial reparations mechanisms, these must be legitimate, accessible, predictable, equitable, transparent and compatible with human rights. The Human Rights Council has asserted, “the remedies that may be obtained from non-State-based grievance mechanisms are usually partial at best, in many cases due to limitations placed on the mechanism’s mandate, available resources, or both.”

In order to ensure the effectiveness of non-judicial reparations mechanisms, these must be legitimate, accessible, predictable, equitable, transparent and compatible with human rights.
IV. The Role of the Judicial Branch in the Internet Era

The advent of the FOB has also triggered debate on the role of this type of non-state council or mechanism, whose legitimacy and operation are far from jurisdictional bodies established in countries governed by rule of law.

Is the FOB a solution to this complex issue? Shouldn’t a decision that determines the suspension of an account of a public official or a person of high public profile enable the possibility of judicial oversight and notify the persons affected? These are some of the questions up for debate.

Based on constitutional tradition and inter-American human rights standards, and in response to conflicting rights that may be impacted with the removal of content affecting freedom of expression in Latin America, the intervention of an independent judge has been firmly guaranteed. Ever since its first reports on freedom of expression on the internet, the Office of the Special Rapporteur for Freedom of Expression at the IACHR has insisted on the need for a court order to take down content. "Specifically, the requirement that intermediaries remove content, as a condition of exemption from liability for an unlawful expression, could be imposed only when ordered by a court or similar authority that operates with sufficient safeguards for independence, autonomy, and impartiality, and that has the capacity to evaluate the rights at stake and offer the necessary assurances to the user."

This point has given rise to constant tension. On one hand, platforms are generally required to comply with court rulings that order content be taken down or deleted, but the volume of decisions to be made by a platform makes a judicial oversight system for all cases almost unfeasible.

Secondly, different specific laws have established forms of conditional liability, such as those related to the protection of intellectual property that order intermediaries to remove reported content and notify the affected party afterwards. These norms have also weakened the idea of judicial oversight on the internet, although they have proven effective at addressing the substantive problem.

Platforms are generally required to comply with court rulings that order content be taken down or deleted, but the volume of decisions to be made by a platform makes a judicial oversight system for all cases almost unfeasible.

On the other hand, faced with the impact of hate speech, platforms have intensified content moderation in an attempt to solve the problem and provide safer spaces, imposing community standards with lower tolerance thresholds. In a recent decision, the Supreme Court of Germany established the invalidity of Facebook’s terms of service with regard to deleting posts and blocking accounts that violate its community standards. The Court based its decision on the fact that the company did not inform the user about the removal of the offensive post – at least after the event – nor provide reasons or offer the right to appeal.

Even though there is consensus of public opinion regarding company reasoning, different voices have warned that the FOB cannot replace judicial mechanisms for decision-making and redress. On the other hand, international experts argue that this type of mechanism should be complementary to judicial guarantees, and that they should coexist in the pursuit of the protection of human rights. Thus, there could be a fruitful dialogue between the decisions of national and international courts and the decisions of mechanisms such as the FOB.
V. Impact on the Regulatory Discussion in Latin America

As mentioned previously, the FOB began to operate in the midst of heated legislative discussion on how to guarantee freedom of expression on the internet.

In this regard, former Rapporteur for Freedom of Expression David Kaye posits the need for dialogue not only between FOB decisions and international law, but also from these decisions towards countries’ internal regulatory agenda. “Decision makers and judges may be watching the Board’s actions and policy recommendations to see where future developments in regulations are heading.”

Indeed, over the past five years there has been increasing legislative debate in Latin America on the need to regulate different aspects of content dissemination on digital platforms. Dozens of bills on the subject have been introduced in Congresses and although they often pursue legitimate ends, the solutions proposed generally include restrictions on freedom of expression or fail to consider how the internet is designed.

Developing legal frameworks to regulate social network content still poses challenges, even in States with high quality democratic institutions. One example is the NetzDG Law enacted in Germany requiring platforms to take down any speech, within 24 hours of receiving a complaint, that supposedly does not comply with the country’s anti-discrimination laws. Platforms’ transparency reports show that the experience in Germany led to the removal of millions of posts, although there are doubts about its real impact on the circulation of hate speech.

In May 2020, the French Parliament enacted a law requiring internet platforms to delete any “manifestly illegal” content within 24 hours under penalty of liability. The “Avia Law,” named after its parliamentary sponsor Laetitia Avia, included several provisions similar to those in the German NetzDG and focused on pedophile and terrorist content. However, before it was enacted, several legislators subjected it to a study by the French Constitutional Council, which in June 2020 declared it unconstitutional.

In countries with greater democratic deficits or who are ruled by authoritarian regimes, this debate is used to justify legislative proposals that often lead to censorship. The most prominent case in the region is Venezuela’s law “Against Hatred, for Tolerance and Peaceful Coexistence,” which established ambiguous terms to force platforms to take down sites and content inciting “hate or violence,” and prison penalties for those who “incite violence” on social media, as well as disproportionate requirements and fines for intermediaries. Indeed, there have been several arrests leading to imprisonment of cyber-activists due to this legislation.

Cuba also enacted a rule through which the State can fine network users by applying vague or ambiguous measures. Indeed, Decree Law 370/2019, which regulates aspects related to the governance of internet infrastructure, also enables authorities to sanction social network users, and forbids any content contrary to “social interest, morality, decency and personal security” to be posted on social media.

A report by the organization Cubalex concluded that at least 28 Cubans, mostly journalists and activists, reported having been penalized for content in their Facebook posts. Fines begin at $120, but in an economic situation like the one in Cuba, in addition to penalizing legitimate speech, the Decree Law acts as a serious deterrent.
In October 2020, Nicaragua enacted its Cybercrime Law, which punishes the “unauthorized dissemination” of “fake information” or information that threatens the population and national security of the country, with up to 10 years of prison. This law, driven by President Daniel Ortega, was rejected by the opposition, who considered it to be an attempt at censorship.

With regard to democratic states in the region, during the past year several regulatory bills were introduced in Brazil that have been introduced to address different aspects of online content and intermediary liability. Some of these bills refer specifically to the political content circulating on the internet, for example Bills 291/2021 and 449/2021 forbid companies on the internet from removing content based on political, ideological or religious criteria. Bills 246/2021, 356/2021 and 388/2021 forbid the interruption of internet services without a prior court order, and establish sanctions against providers who fail to comply. Bill 2630/2020, known as the “Fake News Bill” establishes rules of procedure and transparency for content moderation.

Moreover, following the Trump case in the United States, the Brazilian government introduced a presidential decree modifying the regulations of the Civil Rights Framework for the Internet, so that any exclusion, limitation of posts or suspension of accounts could only be done by means of a court order, with some exceptions. The different initiatives proposed have been criticized by civil society in Brazil because of the risks to users’ freedom of expression.

In Colombia, several bills have sought to address content circulation on social media. The Committee on Transport and Communications is currently studying Bill 215/2020C, the aim of which is to regulate advertising practices on social media. In addition, Bill 048/2020 seeks to establish a series of measures to protect people’s honor and privacy, and provide redress if they are harmed. Bill 600/2021 is also under study, which regulates the responsibility of

With regard to the jurisdictional arena, although the region has a solid inter-American legal framework on freedom of expression, which includes guiding principles, there are still challenges in its application to guarantee free, open, and inclusive internet.

In July 2020, Mexico amended its Federal Copyright Law as part of the negotiations in the Free Trade Agreement with the United States and Canada. The amendment included the requirement for digital platforms to establish notification and removal systems to protect copyright. The system requires Internet Service Providers to delete content when it is reported by the owners of the works involved. If Internet Service Providers do not take measures, they will be held accountable for the infringing content.

The reform was criticized by civil society organizations due to its negative impact on freedom of expression in the country. The organizations noted that the mechanism is a form of prior restraint administrated entirely between private parties with no intervention from the judiciary, in violation of articles 6 and 7 of the Mexican Constitution.
There is a push for regulatory bills with strong state intervention, which do not always respect due process. That is why it is important to ensure that the legislative, administrative and judicial sphere apply minimum standards of protection of free speech.

as well as the American Convention on Human Rights and the International Covenant on Civil and Political Rights. The backlash from this issue led the National Human Rights Commission (CNDH) to file a lawsuit before Mexico’s Supreme Court of Justice challenging the law as unconstitutional because it contained imprecise or ambiguous provisions that violate legal certainty and due process.\(^78\)

In February 2021, a bill was introduced in the Mexican Senate to modify the Federal Telecommunications and Broadcasting Law, establishing the general basis and principles for protecting freedom of expression on social media.\(^79\) The proposal by Senator Ricardo Monreal Ávila endows the Federal Telecommunications Institute with a series of oversight powers and establishes an expedited appeals procedure before the agency on decisions made by “relevant” social media, those with over a million users. It also forbids the use of algorithms or automated technologies to cancel accounts permanently. The initiative caused concern among civil society because its concepts are vague and it opens the door to potential government censorship of dissident voices.\(^80\)

In Peru, the Congressional Committee on Transport and Communications is studying two bills on the liability of online intermediaries. Bill 06383/2020 requires internet operators to install filters to block pornographic content or content related to sexual exploitation of children and adolescents.\(^81\) Bill 07222/2020 seeks to regulate “misuse” of social media and apps, establishing requirements for users, internet providers and web administrators.\(^82\) The bill also establishes the requirement for companies operating social media in Peruvian territory to subscribe to agreements or codes of conduct with the government, in which they accept liability for the content published on their platforms. Civil society criticized the fact that the Bill could infringe upon the network’s freedom of expression and neutrality, and noted the lack of inclusion of multiple stakeholders in the parliamentary debates.\(^83\)

With regard to jurisdiction, although the region has a solid inter-American legal framework on freedom of expression, which includes guiding principles, there are still challenges in its application to guarantee free, open, and inclusive internet. In general, national courts have adopted these principles, but there is still a long way to go, especially due to the diversity of the casuistry represented by the cases related to speech posted on social media which have led to litigation.\(^84\)

This review of legislative initiatives in the region demonstrates the growing interest in addressing the dimensions which are impacted by online content moderation. There is a push for regulatory bills with strong state intervention, which do not always respect due process. That is why it is important to ensure that the legislative, administrative and judicial spheres apply minimum standards of protection of free speech.
The Facebook Oversight Board is a novel mechanism for dealing with increasing issues involved in moderating online content. Through its first decisions, it has generated positive expectations regarding the depth of its legal foundation, incorporation of human rights standards and inclusion of recommendations to the company that go beyond the specific case, with the aim of changing or refining the company’s content moderation policies.

Nevertheless, the mechanism is not exempt from criticism. The “Supreme Court” model is just one of the many forms that could be adopted as non-state moderation mechanisms that are innovative and have enough capacity to influence companies. Moreover, this model may not be the most suitable for platforms with fewer followers and less economic capacity.

To date, only three cases in the region have been taken on by the FOB: two from Brazil and one from Colombia – two of the markets with the most Facebook users in Latin America. Although the decisions included global standards and recommendations, no regional standards have been introduced to date. Civil society organizations in the region make a positive evaluation of current FOB members’ commitment to human rights, but aim to ensure long-term commitment, and they are monitoring the level of company compliance with the decisions.

Following worldwide trends, the Latin American region has introduced legislation seeking to regulate content moderation and speech on social media. Examples so far have shown that this can open the door to potential abuse and violations, so regulators and decision makers should pay special attention to standards and guarantees of rights. In addition, these examples show that decisions made by the FOB in the future, whether or not with reference to the Latin American region, may feed into the region’s regulatory agenda.

After identifying the main issues involved in FOB operation, FOB interaction with different actors, the views of experts in the region, and the regulatory trends in Latin America, we developed the following recommendations, which may serve to guide best practices in online content moderation.
GOVERNMENTS

Respect human rights standards. Several bills have arisen in Latin America seeking to address issues related to the content that circulates internet-based platforms. Even though they may be well-intentioned, many of these bills have democratic deficiencies or are not aligned with global and regional standards on freedom of expression. When content moderation regulations are discussed, members of congress and decision makers should adhere to human rights standards as well as consider the internet’s decentralized model of communication.

Make decisions based on due legal process. When a State needs to make a decision to restrict, block or suppress content that circulates on an internet platform, it should wait for an order issued by an independent judiciary body acting under the rules of due process and capable of applying the test of legitimacy, necessity and proportionality established by international law. Similarly, the state, in particular the judicial branch, holds a position as guarantor to users who may be affected by private sector decisions and wish to resort to judicial review. For such purpose, it is necessary to educate decision-makers, members of congress, judges and other actors such as prosecutors and advocates.

Promote multisectoral settings. Promoting self-regulation models or Social Media Councils made up of multiple stakeholders may be one way to help platforms improve their systems. As recommended by the Special Rapporteurs for Freedom of Expression, these Councils contribute to better decision-making processes on internet content, and may be used to review cases or as advisors, but without replacing private or state initiatives.

Analyze FOB decisions and other self-regulation mechanisms. The decisions adopted so far by the FOB have proven to have a high impact on discussions about online freedom of expression, in addition to including in-depth consideration of human rights in their arguments. Notwithstanding the fact that states are not bound by the decisions of the FOB, countries in the region should pay attention to the selected cases, the decisions made on those cases and the policy recommendations made to Facebook. Analysis of these recommendations together with the responses of the company can provide suggestions and guidance regarding future regulation of large internet platforms.
FACEBOOK OVERSIGHT BOARD

Incorporate regional standards. FOB decisions to date have included the international framework on human rights. It is noteworthy that the FOB has resorted to global covenants, compacts and soft law instruments on the subject. For future decisions, the Board should also include regional standards for the protection of freedom of expression. The American Convention on Human Rights, derived from Inter-American Commission on Human Rights reports and Inter-American Court case law can provide broad guarantees for stakeholders in the region.

Implement clear, contextualized consultation mechanisms. Establish clear mechanisms for consulting organizations and experts in the jurisdictions where it operates, regarding national and international standards, thereby improving transparency. This will help improve considerations on the local, political and social context in the decisions made.

Expand the range of jurisdictions served. The Board’s first decisions have mainly focused on jurisdictions relevant to Facebook due to the number of users present in those countries. The processes for selecting cases should consider not only jurisdictions with large numbers of users, but also a range of different countries from each region, given that the platform may have outsize impact on the public forum in countries of less commercial importance.

Expand complaint and interaction mechanisms. The FOB Bylaws provide that only Facebook or Instagram users can file complaints through this mechanism. This leaves out non-users who may be directly involved or harmed by content and the decisions made with regard to such content. Similarly, international bodies such as the Rapporteurships that promote and protect different rights of the regional systems and the universal system should have a direct channel for contributing to decisions and recommendations made by the FOB on moderation policy.

Establish accountability mechanisms. Improve levels of transparency and detail in the follow-up reports on platform compliance with recommendations, including answers to FOB requests for information.
PLATEFORMS

Undertake long-term commitments to transparency. Platform companies must adopt a long-term commitment to the transparency of their operations, as well as the remedies they adopt in specific cases and that can avoid harm to users or groups of users. It is particularly important to provide relevant information to the academic sector and civil society organizations that conduct research on digital rights.

Comply with international standards. In creating these new mechanisms, companies must specifically ensure the independence and transparency of members and operations. Therefore, questions of due process and accountability cannot be neglected. Some of these instruments are the United Nations Guiding Principles on Business and Human Rights and the joint declarations of the Rapporteurs on Freedom of Expression of different regional and international bodies.

Explore new mechanisms. The experience of Facebook and the FOB is just one of the many forms that can be adopted as external review mechanisms of the decisions made by internet platforms. The industry is not expected to be limited to the model and functional structures created by Facebook. In this regard, other multiple stakeholder councils or action guidelines for content moderation based on the principles of freedom of expression and in the framework of human rights may be some initiatives for the industry to explore.

Commit to the decisions and recommendations. Once an independent mechanism has been established – as in the case of the FOB analyzed herein – companies need to respect its decisions and take steps to comply with the recommendations. Generating a virtuous circle among case selection, independent review and the adoption of corrective measures for some policies – in addition to redress for the specific case – can lead to substantive improvement of the previous problems involved in content moderation.

Engage in dialogue with other companies. The challenges of content moderation should be understood as broader, shared problems that go beyond each company’s business logic. Each experience may contribute lessons to be learned and adapted to the needs of other platforms. Companies that operate social media platforms should engage in dialogue with each other about the need to establish independent mechanisms to review content moderation policies and to follow up on their outcomes. In this regard, companies should promote and support decisions adopted by other external review mechanisms when they are consistent with international human rights standards.
Continue the exchange with the FOB. Since the beginning of the FOB’s development, civil society in Latin America has closely followed and actively participated in the consultations of the new mechanism. To maintain the relevance of the region in the FOB’s decisions, organizations need to continue to send comments, report on the context of the decisions and conduct critical analyses on the recommendations and compliance with them.

Foster greater participation. In addition, organizations involved in digital rights and protection of online freedom of expression should disseminate information on the FOB and encourage participation of other organizations who have the capacity to analyze the impact of content on human rights. These organizations increasingly use Facebook’s platforms, so they may be directly affected by company decisions. Disclosing mechanisms to appeal internally and before the FOB may foster their participation and better decision-making.

Monitor government initiatives. The survey of national legislation in several countries in the region on online speech reveals certain distancing from democratic standards. That is why it is essential that civil society organizations maintain their role of monitoring these initiatives, and securing support from other organizations in the region.
ENDNOTES


10 United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-oper-

11 Facebook is a member of the Global Network Initiative, a network of multiple stakeholders that promote laws and policies that protect and respect freedom of expression and privacy rights throughout the world. See more at https://globalnet-workinitiative.org.


33 Inter-American Dialogue, International Justice Clinic at the University of California, Irvine School of Law and Columbia Global Freedom of Expression at Columbia University organized the event "The Decisions of Facebook's Oversight Board: Implications for the Global South, particularly in Latin America." Written summary of the event available at https://www.thedialogue.org/analysis/the-decisions-of-facebook-oversight-board-implications-for-the-global-south-particularly-in-latin-america/.


65 Caroline de Toni, “Dispõe sobre a responsabilidade civil de provedores de aplicações de internet pela atividade de moderação, na forma de rotulagem de conteúdo que expresse a opinião de usuário, e assim caracterize exercício de liberdade de expressão, comunicação e manifestação de pensamento,” Câmara Dos Deputados, 2021, https://www.camara.leg.br/proposicoesWeb/fichadetratamitacao?idProposicao=2269422.


77 R3D, “NiCensuraNiCandados: Condenamos la aprobación de reformas que establecen mecanismos de censura en Internet y criminalizan la elusión de candados digitales,” R3D: Red En Defensa de Los Derechos Digitales, July 1, 2020, https://r3d.mx/2020/07/01/nicensuranicandados-condenamos-la-aprobacion-de-reformas-que-establecen-mecanismos-de-censura-en-internet-y-criminalizan-la-elusion-de-candados-digi-...


83 “Preocupación en Perú por proceso del debate y contenidos de proyectos de Ley sobre regulación de Internet”, Observacom, April 30, 2021, https://www.observacom.org/preocupacion-en-peru-por-proceso-del-debate-y-contenidos-de-proyectos-de-ley-sobre-regulacion-de-internet/.

About the Peter D. Bell Rule of Law Program

Established in 2015 with support from the Ford Foundation and named in honor of a founding Dialogue co-chair, the Peter D. Bell Rule of Law Program aims to elevate policy discussions around democracy and human rights, corruption and transparency, and citizens security in the Americas.