

MLRC Media Law Conference  
October 4-6, 2023  
Lansdowne Resort Hotel, Leesburg, VA

PLENARY 4 Oct. 6<sup>th</sup> 9:30 – 10:30am

Online Content Moderation: The future of online content moderation and the impact of the Oversight Board, an independent external body created to help Meta answer some of the most difficult questions around freedom of expression for Facebook and Instagram

**Brent Harris** is an American lawyer and technology executive. As vice president of the augmented reality family at Meta Platforms, Harris is focused on how we build the next mobile computing platform for people and businesses across the world. Previously, Harris was the vice president of governance at Meta. He pioneered new ways to govern artificial intelligence and the Metaverse, and he led the company's work to build the Oversight Board for Facebook and Instagram. Harris holds bachelor's, master's, and law degrees from Stanford University, where he edited the *Stanford Law Review*. His writing has appeared in op-eds for the *New York Times*, and recently he was a visiting scholar at the Stanford Graduate School of Business. Originally from Oklahoma, he lives in California with his wife and son.

**Alison Stein** is Co-Chair of the Content, Media, and Entertainment and Technology practices at Jenner & Block. Alison represents technology companies, media and entertainment organizations, video game companies, and television and movie studios. Clients rely on Alison's proven counsel and strategic know-how to address issues ranging from content moderation, copyright law, content protection, generative AI and other immersive technologies, to trademark, trade secrets, compliance, defamation, and contractual and licensing disputes. Alison serves as a Co-Chair of the Media Law Resource Center's Internet & Technology Law Committee, and on the Board of Directors for Volunteer Lawyers for the Arts. Alison is frequently named among the top lawyers in her field by *Chambers*, was named a "Notable Woman in the Law" by *Crain's*, and recognized in both 2022 and 2023 as a "Power Lawyer" by *The Hollywood Reporter*. Alison has worked closely with Brent Harris the governance team at Meta for many years including on the founding of the Oversight Board.

## **Issues include:**

- What was the impetus for the Oversight Board, how was it created and implemented?
- How independent is the Board? Is independence necessary for legitimacy?
- What are the sources for the Oversight Board decisions? How do cases get to them?
- How can international human rights standards be applied to commitments to maintain online safety and security?
- Is the Oversight Board a model for other social media self-regulation?
- What is the future of online content moderation and the impact of AI? Augmented reality?

## **Academic Commentary**

Price, Monroe; Price, Joshua: Building Legitimacy in the Absence of the State: Reflections on the Facebook Oversight Board (2023). USC Anneberg International Journal of Communication.

<https://ijoc.org/index.php/ijoc/article/view/19915/4171>

Meta created a Facebook Oversight Board to burnish or attempt to burnish the legitimacy of its social media platforms. The Oversight Board can compensate for regulatory failure, particularly deficits occasioned by the

absence of explicit direct government supervision. These deficits are usually marked by significant public criticism of platforms for their perceived failure to moderate content adequately. In a global context in which the interplay between formal regulation and platform-initiated efforts to moderate content becomes ever more intense and heated, innovations, like the work of the Board, can be enlightening. We look at aspects of the Board's creation to identify steps thought to manufacture legitimacy. In this case and others, the platform in conversation with stakeholders yields additional opportunities for review of decisions to take down or retain material and to allow additional perspectives on policy issues. In the process of review, we identify ways that platforms generate attributes of legitimacy, enhance strategic narratives, and develop an epistemic community—all with the goal of increasing legitimacy.

Douek, Evelyn, Facebook's 'Oversight Board:' Move Fast with Stable Infrastructure and Humility (April 4, 2019). 21 N.C. J. L. & Tech. 1 (2019), Available at SSRN: <https://ssrn.com/abstract=3365358>

Facebook's proposed Oversight Board is one of the most ambitious constitution-making projects of the modern era. With pre-existing governance of tech platforms delegitimized by the ongoing "techlash," this represents a pivotal moment when new constitutional forms can emerge that will shape the future of online discourse. For all the potential of the Facebook Oversight Board, there are many things it cannot be. It will not hear a sufficient amount of cases to be a meaningful response to calls for greater due process in individual content moderation decisions. Nor will it be able to become a font for global speech norms for the worldwide platform. The true value that the Board can bring to Facebook's content moderation ecosystem lies between these two extremes of individual error correction and the settlement of globally applicable speech rules. The institutional offering of the Board should focus on two primary, but more modest, functions. First, it can help highlight weaknesses in the policy formation process at Facebook, removing blockages (such as blind spots and inertia) in the "legislative process" leading to the formulation of its Community Standards. Second, by providing an independent forum for the discussion of disputed content moderation decisions, the Board can be an exemplar of the public reasoning necessary for persons in a pluralistic community to come to accept the rules that govern them, even if they disagree with the substance of those rules. Understanding the institutional

role of the Board in these terms provides useful insights into the institutional design that will best help it achieve these goals.

Klonick, Kate, *The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression* (June 30, 2020). *Yale Law Journal*, Vol. 129, No. 2418, 2020, Available at SSRN: <https://ssrn.com/abstract=3639234>

For a decade and a half, Facebook has dominated the landscape of digital social networks, becoming one of the most powerful arbiters of online speech. Twenty-four hours a day, seven days a week, over two billion users leverage the platform to post, share, discuss, react to, and access content from all over the globe. Through a system of semipublic rules called “Community Standards,” Facebook has created a body of “laws” and a system of governance that dictate what users may say on the platform. In recent years, as this intricately built system to dispatch the company’s immense private power over the public right of speech has become more visible, Facebook has experienced intense pressure to become more accountable, transparent, and democratic, not only in how it creates its fundamental policies for speech but also in how it enforces them.

In November 2018, after years of entreaty from the press, advocacy groups, and users, CEO and founder Mark Zuckerberg announced that Facebook would construct an independent oversight body to be researched, created, and launched within the year. The express purpose of this body was to serve as an appellate review system for user content and to make content-moderation policy recommendations to Facebook. This Feature empirically documents the creation of this institution, now called the Facebook Oversight Board. The Board is a historic endeavor both in scope and scale.

The Feature traces the events and influences that led to Facebook’s decision to create the Oversight Board. It details the year-long process of creating the Board, relying on hundreds of hours of interviews and embedded research with the Governance Team charged with researching, planning, and building this new institution.

The creation of the Oversight Board and its aims are a novel articulation of internet governance. This Feature illuminates the future implications of the new institution for global freedom of expression. Using the lens of adjudication, it analyzes what the Board is, what the Board means to users,

and what the Board means for industry and governments. Ultimately, the Feature concludes that the Facebook Oversight Board has great potential to set new precedent for user participation in private platforms' governance and a user right to procedure in content moderation.

Arun, Chinmayi, Facebook's Faces (March 15, 2021). Forthcoming Harvard Law Review Forum Volume 135 , Available at SSRN:

<https://ssrn.com/abstract=3805210> or <http://dx.doi.org/10.2139/ssrn.3805210>

The Facebook Oversight Board's decision about the suspension of Donald Trump's account is different from the Board's other cases because it interests states. The 'Trump Ban' case affects the Board's reputation and Facebook's relationships with states and publics. We will not understand the case's impact if we do not understand these relationships.

Scholarship about social media platforms discusses their relationship with states and users. The Essay is the first to expand this theorization to account for differences among states, the varying influence of different publics and the internal complexity of companies. Theorizing Facebook's relationships this way includes less influential states and publics that are otherwise obscured. It reveals that Facebook engages with states and publics through multiple, parallel regulatory conversations, further complicated by the fact that Facebook itself is not a monolith. This Essay argues that Facebook has many faces – different teams working towards different goals, and engaging with different ministries, institutions, scholars and civil society organizations. Content moderation exists within this eco-system.

This Essay's account of Facebook's faces and relationships shows that less influential publics can influence the company through strategic alliances with strong publics or powerful states. It also suggests that Facebook's carelessness with a seemingly weak state or a group, may affect its relationship with a strong public or state that cares about the outcome.

To be seen as independent and legitimate, the Oversight Board needs to show its willingness to curtail Facebook's flexibility in its engagement with political leaders where there is a real risk of harm. This Essay hopes to show Facebook that the short-term retaliation from some states may be balanced out by the long-term reputational gains with powerful publics and powerful

states who may appreciate its willingness to set profit-making goals aside to follow the Oversight Board's recommendations.

Bietti, Elettra, A Genealogy of Digital Platform Regulation (June 3, 2021). 7 Geo. L. Tech. Rev. 1 (2023), Available at SSRN: <https://ssrn.com/abstract=3859487> or <http://dx.doi.org/10.2139/ssrn.3859487>

I connect current efforts to 1990s debates around the regulation of cyberspace: contestations on the meaning of freedom, law, power, and democracy in digital spaces. I isolate three paradigmatic views, or moments, in early Internet regulation discourse: anarcho-libertarian, liberal, and critical views. I ask how these three views or moments have shaped and led to a similar spectrum of three views on how to regulate digital platforms and promote freedom in digital spaces: libertarian aversion to regulation; liberal perspectives on self-regulation, fiduciary obligations, data protection, competition, and utility regulation; and critical accounts of platform governance.

The move from an Internet of networks to an Internet of platforms represents a significant shift: from a hybrid, decentralized environment where freedom seemed the norm, to a centralized space where the default is privatized enclosure. Still, 1990s and current understandings of digital freedom, power, and law are pervaded by similar market-liberal path-dependencies that continue to facilitate the consolidation of private power in digital environments. I suggest two steps towards a post-neoliberal approach to digital policy.

## **Selected Oversight Board Decisions**

United States posts discussing abortion (2023)  
The Oversight Board has overturned Meta's original decisions to remove three posts discussing abortion and containing rhetorical uses of violent language as a figure of speech.

<https://www.oversightboard.com/decision/IG-FZSE6J9C/>

The Oversight Board overturns Meta’s original decisions to remove two Facebook posts and one Instagram post, all of which discussed abortion. The Board finds that the three posts did not violate Meta’s Violence and Incitement policy, as they did not incite or threaten violence but were rhetorical comments about abortion policy. Meta has acknowledged that its original decisions were wrong, and that the content did not violate its Violence and Incitement policy. The Board selected these cases to examine the difficult content moderation problem of dealing with violent rhetoric when used as a figure of speech as well as its potential impact on political expression....

Discussion of abortion policy is often highly charged and can include threats that are prohibited by Meta. Therefore, it is important Meta ensure that its systems can reliably distinguish between threats and non-violating, rhetorical uses of violent language.

Since none of these cases are ambiguous, the errors suggest there is scope for improvement in Meta’s enforcement processes. While such errors may limit expression in individual cases, they also create cyclical patterns of censorship through repeated mistakes and biases that arise from machine-learning models trained on present-day abusive content. Additionally, these cases show that mistakenly removing content that does not violate Meta’s rules can disrupt political debates over the most divisive issues in a country, thereby complicating a path out of division.

Meta has not provided the Board with sufficient assurance that the errors in these cases are outliers, rather than being representative of a systemic pattern of inaccuracies.

#### Promoting Ketamine for non-FDA approved treatments (2023)

The Oversight Board has overturned Meta’s decision to leave up a user’s Instagram post discussing their experience using ketamine as a treatment for anxiety and depression.

<https://www.oversightboard.com/decision/IG-TOM6IXVH/>

As the content in this case was part of a “paid partnership,” clearly promoted the use of ketamine, and was not covered by an exception, it violated these policies. In response to the Board’s questions, Meta acknowledged that not all content with a “paid partnership” label is reviewed against its Branded

Content policies, that moderators reviewing content at scale cannot see this label, and that they cannot reroute content to the specialized team in charge of enforcing the Branded Content policies. This greatly increases the risk of under-enforcement against this kind of content. As such, the Board urges Meta to ensure that it reviews content against all relevant policies, including its Branded Content policies.

#### Cambodian prime minister (2023)

The Oversight Board has overturned Meta’s decision to leave up a video on Facebook in which Cambodian Prime Minister Hun Sen threatens his political opponents with violence.

<https://www.oversightboard.com/decision/FB-6OKJPNS3/>

The Board finds that Meta was wrong to apply a newsworthiness allowance in this case, as the harm caused by allowing the content on the platform outweighs the post’s public interest value. Given Hun Sen’s reach on social media, allowing this kind of expression on Facebook enables his threats to spread more broadly. It also results in Meta’s platforms contributing to these harms by amplifying the threats and resulting intimidation.

#### Reaction

Meta rejects own board’s request to suspend account of Cambodian strongman  
Washington Post August 30, 2023

Meta, which owns Facebook, Instagram and WhatsApp, has struggled in recent years to balance freedom of expression with abusive online behavior. The company said it based the Hun Sen decision on protocol governing the speech of public figures that it developed after its suspension of U.S. President Donald Trump over incendiary posts that he made during the Jan. 6, 2021, Capitol riots. In its analysis, the company said, it determined that Hun Sen’s video did not occur in what it considered a crisis situation and therefore that a suspension was unwarranted. The Cambodian government said in a statement Tuesday that the company had exercised “fair judgment” and that its decision “confirms the integrity” of content on Hun Sen’s page. It added that Meta representatives were welcome to continue to work in the country, whereas Oversight Board members were still considered *personae non grata*.

<https://www.washingtonpost.com/world/2023/08/30/meta-cambodia-facebook-hun-sen/>



### Brazilian general's speech (2023)

The Oversight Board overturns Meta's original decision to leave up a Facebook video featuring a Brazilian general calling people to "go to the National Congress and the Supreme Court."

<https://www.oversightboard.com/decision/FB-659EAWI8/>

The Board is deeply concerned that despite the civil unrest in Brazil at the time the content was posted, and the widespread proliferation of similar content in the weeks and months ahead of the January 8 riots, Meta's content moderators repeatedly assessed this content as non-violating and failed to escalate it for further review. In addition, when the Board asked Meta for information on specific election-related claims on its platforms before, during, and after the Brazilian elections, the company explained that it does not have data on the prevalence of such claims. The content in this case was finally removed more than two weeks later, by which point the violating event it called had already occurred, and only after the Board brought the case to Meta's attention.

### Former President Trump's suspension (2021)

The Board has upheld Facebook's decision, on 7 January 2021, to restrict then-President Donald Trump's access to posting content on his Facebook Page and Instagram account.

<https://www.oversightboard.com/decision/FB-691QAMHJ/>

In this case, Facebook asked the Board to answer two questions:

Considering Facebook's values, specifically its commitment to voice and safety, did it correctly decide on January 7, 2021, to prohibit Donald J. Trump's access to posting content on Facebook and Instagram for an indefinite amount of time?

In addition to the board's determination on whether to uphold or overturn the indefinite suspension, Facebook welcomes observations or recommendations from the board about suspensions when the user is a political leader.

#### 1. Decision summary

The Board upholds Facebook’s decision on January 7, 2021, to restrict then-President Donald Trump’s access to posting content on his Facebook page and Instagram account.

However, it was not appropriate for Facebook to impose the indeterminate and standardless penalty of indefinite suspension. Facebook’s normal penalties include removing the violating content, imposing a time-bound period of suspension, or permanently disabling the page and account.

The Board insists that Facebook review this matter to determine and justify a proportionate response that is consistent with the rules that are applied to other users of its platform. Facebook must complete its review of this matter within six months of the date of this decision. The Board also makes policy recommendations for Facebook to implement in developing clear, necessary, and proportionate policies that promote public safety and respect freedom of expression.

#### 4. Relevant standards

Under the Oversight Board’s Charter, it must consider all cases in light of the following standards:

##### I. Facebook’s content policies:

.... Facebook’s Community Standard on Violence and Incitement states it “remove[s] content, disable[s] accounts, and work[s] with law enforcement when [it] believe[s] there is a genuine risk of physical harm or direct threats to public safety.” The Standard specifically prohibits: “Statements advocating for high-severity violence” and “Any content containing statements of intent, calls for action, conditional or aspirational statements, or advocating for violence due to voting, voter registration or the administration or outcome of an election.” It also prohibits “Misinformation and unverifiable rumors that contribute to the risk of imminent violence or physical harm.”

##### II. Facebook’s values:

Facebook has five values outlined in the introduction to the Community Standards which it claims guide what is allowed on its platforms. Three of these values are “Voice,” “Safety,” and “Dignity.”

Facebook describes “Voice” as wanting “people to be able to talk openly about the issues that matter to them, even if some may disagree or find them objectionable. [...] Our commitment to expression is paramount, but we recognize that the Internet creates new and increased opportunities for abuse.”

Facebook describes “Safety” as Facebook’s commitment to “mak[e] Facebook a safe place” and states that “Expression that threatens people has the potential to intimidate, exclude or silence others and isn’t allowed on Facebook.”

Facebook describes “Dignity” as its belief that “all people are equal in dignity and rights” and states that it “expect[s] that people will respect the dignity of others and not harass or degrade others.”

### III. Human rights standards:

... The Oversight Board is called to evaluate Facebook’s decision in view of international human rights standards as applicable to Facebook. The Board analyzed Facebook’s human rights responsibilities in this case by considering human rights standards including: The right to freedom of expression: International Covenant on Civil and Political Rights ( ICCPR), Articles 19 and 20; as interpreted in General Comment No. 34, Human Rights Committee (2011) ( General Comment 34); the Rabat Plan of Action, OHCHR, (2012); UN Special Rapporteur on freedom of opinion and expression report A/HRC/38/35 (2018); Joint Statement of international freedom of expression monitors on COVID-19 (March, 2020).

## Reactions

The Oversight Board’s Trump Decision Highlights Problems with Facebook’s Practices

ACLU (May 2021)

Facebook’s initial decision to suspend Trump’s account for a defined and limited time, and the OB’s decision to uphold it, is understandable in light of the events of Jan. 6 and Trump’s part in spreading outright lies about the electoral process in the weeks and days leading up to those events. But the rule Facebook claimed to apply here — its community standard prohibiting

“praise and support of dangerous individuals and organizations” — is too vague, and its application in this case offers little clarity.

<https://www.aclu.org/news/free-speech/the-oversight-boards-trump-decision-highlights-problems-with-facebooks-practices>

Facebook Oversight Board Affirms Trump Suspension - For Now  
EFF (May 2021)

Today’s decision affirms, once again, that no amount of “oversight” can fix the underlying problem. First, while the Oversight Board rightly refused to make special rules for politicians, rules we have previously opposed, it did endorse special rules and procedures for “influential users” and newsworthy posts. These rules recognize that some users can cause greater harm than others. On a practical level, every decision to remove a post or suspend an account is highly contextual and requires often highly specific cultural competency. But we agree that special rules for influential users or highly newsworthy content requires even greater transparency and the investment of substantial resources.

<https://www.eff.org/deeplinks/2021/05/facebook-oversight-board-affirms-trump-suspension-now>