Crisis - Meta Oversight Board
Background Guide

Georgia Tech Model United Nations

the 24th GTMUN
High School Conference
October 9th-10th, 2023

Oversight Board

Crisis Committee
Meta Oversight Board

Introduction

Free speech is a right protected in a large majority of the world, with numerous legal precedents creating guidelines and cases around its use. These guidelines, though, did not carry over seamlessly to the internet, following its creation. The most prominent laws in this space are primarily western, such as the American Section 230. Section 230 was a byproduct of several defamation cases filed in the early 1990's and had the effect of indemnifying website operators from any user-posted content even if they engaged in moderation. With only country-specific laws to guide them, social media platforms have created their own guidelines on speech that aim to fill gaps left by inconsistent regulation, while still conforming to the 193 separate regulatory regimes globally. Moreso than the rules themselves, serious questions have been raised as to the impact of concentrating power over speech in the hands of unelected corporate leaders. Indeed, 94% of social media use occurs on just four sites, owned by three companies.[1]

[1] Statista, "Leading social media websites in the United States as of March 2023, based on share of visits"

Without the legal ability to adjudge cases surrounding speech, moderation decisions were kept entirely in-house without any external review or appeal. After significant controversy, Meta in 2018 became the first platform to create a body that handles appeals and functions as a supreme court with regards to all Meta moderation decisions. Moderation decisions Meta takes typically include content takedowns, limiting content promotion, or banning users who have posted such content before. This board was ideated by Meta CEO Mark Zuckerberg, who also approved the first charter of the body. Due to the relationship of the board and meta itself, there have been significant questions surrounding the nature of decisions by the board -- specifically if they are binding or not. Facebook has already overturned one board recommendation, and has no legal obligation to follow further rulings.[2]

Fig 1. Oversight Board spokesman Dex Hunter-Torricke and member Alan Rusbridger at an Axios media event

With the conflict between rights against government infringement of speech and the privately-enacted moderation of social media platforms, moderation has emerged as a fault line in broader society. Moderation was even cited as the product of social media companies by tech columnist Nilay Patel (source), cementing perception of moderation as a key facet of any platform. More specifically, the issue has come down on political fault lines globally, with some claiming that platforms have acted to suppress unfavorable speech in service of a broader agenda. Others have seen the issue as a problem of community moderation, keeping the online platforms billions of people use safe and free from hate speech.[3]

At the intersection of these two points of view sits the Meta oversight board, straddling the public and private regulations that govern speech. In imposing a legal structure on a private business, Meta has become the executioner of the board's decisions, relinquishing to it the roles of judge and jury. But the real question is not who inside Meta should make those decisions, rather it is should Meta have the powers to make those decisions in the first place? The owners of social media platforms are largely people of immense wealth who were not elected to their roles -- should they be the ones to determine what the masses can post? Changing laws related to online content has been frequently mentioned as a method of keeping moderation in line with the preferences of governments, but this may be no better, as governments globally would greatly enjoy the ability to control what their citizens discuss online. Balancing these interests, both far from the ideal, is what the oversight board and the broader content moderation ecosystem are tasked with.

[2] Cecilia Kang, "What Is the Facebook Oversight Board?"

[3] Nillay Patel, "Welcome to hell, Elon"

History

To understand the current context of moderation, it is important to consider the historical origins of free speech and how they have shaped the modern understanding of that right. In a legal sense, the right was first established for members of the British parliament while lawmaking in 1689 through the Bill of Rights of 1689.[4] This right gave members of parliament legal protection for anything said during a formal session of parliament, to protect the proposal of unfavorable ideas in debate. This confirms the view that freedom of speech was seen of some import, but that import was practical use. Communication, not persuasion was the goal of speech. With the expansion of the right to free speech during the french revolution, the responsibility to bear the abuses of that right was added. This brought into consideration the negative impacts of the right and codified in law another key tenet of the current understanding of the right. This responsibility is something that has frequently appeared in recent history, with some claiming that removal from a social media platform is the speaker taking responsibility for their speech.

On the internet, the delineation between who is speaking and who is not was the first point of debate. If an incendiary comment is left on a website, is the owner of the website responsible, or is the person who posted the comment? This question was hugely important for liability cases in the vein of Sullivan v. New York Times, as making websites liable for all content posted would have been ruinous for most website operators. Section 230 of the CDA answered this question by making websites not liable for user-posted content on their websites, even if they moderate that content to their own community standards. Section 230 has often been called the bedrock of the modern internet, as without it, many websites would be endlessly sued for content users post or create.

The most modern concept of freedom of speech and the internet is the one established 27 years ago, with piecemeal updates by legislators globally. In the US, most regulation has centered around limiting the moderation of content on a state-by-state basis. Consider the 2022 law passed in Texas that prevents "viewpoint discrimination."[5] This law is a direct response to the perceived heavy hand that corporations take with certain content. This law also reflects the nature of free speech as a negative right, a right against government influence. The inverse of this law, one that requires the removal of content, could not pass in a country with legal protection of speech. Limits on the right to free speech could only ever happen in line with strong public opinion and could only cover the most extreme of content. That is the approach taken in several European countries, notably Germany, which bans certain content related to the holocaust.[6] In certain contexts, certain popul-

[4] United Kingdom Parliament, "Bill of Rights 1689"

[5] Jesus Vidales, "Texas social media "censorship" law goes into effect after federal court lifts block"

[6] Human Rights Watch, "Germany: Flawed Social Media Law"

ations may see this infringement as a justified limit while in others it would be viewed as a major limitation.
Moderation, then, is expected to conform to the common understanding of freedom of speech and social expectations. On the Meta platforms, these rules are the terms and conditions of Meta, enforced by the platform and ruled on by the Oversight board. Without a strong board, these regulations could be seen as arbitrary, but a too powerful board would concentrate power over those rules to a problematic degree.

Current

The current situation, as previously stated, is defined along several major fault lines. Specific countries may have their own moderation policies and certain political groups may have their own views on those policies.

Moderation has been frequently used as a tool of repressive governments to prevent certain types of speech from making their way to the public. Consider the extensive moderation network of the People's Republic of China, with anti-government content rapidly removed. The response of most social media platforms has been to simply ignore the chinese market due to the legal restrictions. While this may be the easiest approach, is it the right one? Would a heavily censored version of social media be better than no social media? The repression of content takes many forms, including the requirements of certain governments that platforms have local workers in their countries, presumably to allow governments to pressure platforms by threatening to arrest their employees. In these cases, should an overview body order that changes be made to moderation decisions to protect employees? Or should freedom of speech come before individual safety?

The political fault lines of content moderation have frequently come up, given that attention on moderation is a product of the American election of 2016 and its aftereffects. Right-wing groups often claim to have been banned or subject to reduced visibility due to their political beliefs, especially figures on the extreme right. While many see these views as hateful, a large contingent sees them as valid political positions. Would reduction in visibility of content disproportionately created by members of one political group be valid, even if it reduces the visibility of their political views?

The ability to moderate content has also been a sticking point for current moderation efforts. Certain languages and regions severely lack translators, making moderation of that content difficult or impossible.[7] In India, this has allowed political violence against the Muslim minority to occur

[17] The Washington Post,"How Facebook neglected the rest of the world, fueling hate speech and violence in India"

and multiple people have lost their lives as a result. With this, can moderation of content on a social media platform be considered a responsibility of those who operate it? This question also applied to certain social media platforms that market themselves as moderation-free to appeal to certain political groups. By allowing users to avoid consequences for their speech that they would face on other platforms, these few platforms have been able to attract a healthy user base. There has been little consideration of the impact that allowing speech to silo off into separate groups has had on the understanding of the right to free speech as well.

The siloing of content into discrete groups has also been observed within platforms themselves. Content algorithms have proven themselves to guide users into certain content -- this siloing within platforms themselves has created a multitude of universes of speech.[8] The Meta oversight council, then, is taken with overseeing multiple universes of speech across multiple platforms. The board cannot conceivably hear every case, and there is a lack of authority within the zone that the board does not oversee, comparable to a scenario where only a supreme court and local court-house exist, with little appellate bandwidth.

Beyond just oversight councils and other"official" bodies, certain social media platforms have allowed users to moderate content on their sites. The community notes feature on twitter is the most prominent example of this, with users being able to add context to tweets deemed to be misinformation.[9] This feature does allow users to take on the more official duty of moderation and dealing with misinformation, and can help to bridge the gap of inadequate local translation efforts in certain regions. Features such as this, though, do run the risk of preventing certain content from being adequately seen by the audience it was meant for. Ongoing efforts by the PRC to limit negative overseas discussion of it and its policies is one risk that a feature like this must take into account.

[8] Brookings,"Echo chambers, rabbit holes, and ideological bias: How YouTube recommends content to real users?"

[3] Twitter Help Center, "About Community Notes on Twitter"

Directives

With the questions discussed before, the question now comes to what can be done by the oversight board in this global climate. The chair would look favorably upon discussion of the following topics, into which delegates are encouraged to conduct further research.

- Can one body conceivably apply one set of standards across the entire world?
  - If there is one set of standards, who decides on it?
- Is it right for the power over modern communication platforms to rest with one group in particular? How should that power be distributed?
- How should content moderation function when a language barrier exists?
- What should the policy around authoritarian regimes and moderation be?
- What does the future of content moderation look like? Community-based? Top-down?

These questions are merely starting points from the mind of the crisis director. Delegates are encouraged to investigate other potential questions to answer, and to be very flexible during the conference itself.

Work Cited

Isaac, Mike. "What Is the Facebook Oversight Board?" The New York Times, 5 May 2021, https://www.nytimes.com/2021/05/05/technology/What-Is-the-Facebook-Oversight-Board.html.

Boggs, Austin. "Elon Musk's Twitter acquisition has been a speech moderation disaster." The Verge, 28 Oct. 2022, https://www.theverge.com/2022/10/28/23428132/elon-musk-twitter-acquisition-problems-speech-moderation.

"Bill of Rights 1689." UK Parliament, https://www.parliament.uk/about/living-heritage/evolutionofparliament/parliamentaryauthority/revolution/collections1/collections-glorious-revolution/billofrights/.

McCullough, Jolie. "Federal judge blocks Texas social media law that would let residents sue over censorship." Texas Tribune, 16 Sept. 2022, https://www.texastribune.org/2022/09/16/texas-social-media-law/.

"Germany: Flawed Social Media Law." Human Rights Watch, 14 Feb. 2018, https://www.hrw.org/news/2018/02/14/germany-flawed-social-media-law

Toor, Amar. "Facebook is a major vector for hate speech and misinformation in India, report finds." The Washington Post, 24 Oct. 2021, https://www.washingtonpost.com/technology/2021/10/24/india-facebook-misinformation-hate-speech/.

West, Darrell M., et al. "Echo chambers, rabbit holes, and the challenges of ideological diversity." Brookings Institution, 17 May 2018, https://www.brookings.edu/research/echo-chambers-rabbit-holes-and-ideological-bias-how-youtube-recommends-content-to-real-users/.

"Community notes." Twitter Help Center, https://help.twitter.com/en/using-twitter/community-notes.