

Essay 2

Prompt: Describe the policies and processes that Facebook uses to moderate content on its site that are discussed in readings and in class. Suggest how the policy or process could be improved. What motivates your suggested improvement? Why do you think this improvement hasn't been adopted already?

As the digital age and social media continue to dominate the social commentary of our everyday lives, one of the problems that arises with anonymity and free speech is the need to moderate content. U.S. law approves of the freedom of speech, the idealistic notion proclaimed by Voltaire that one has the right to say whatever they want. However, the limitations of this statement have been tested like never before by online communities and platforms. The shroud of privacy that legally comes with online presence, while protecting user information, further enables those with more malicious intents. One prime example of an institution facing this dilemma is Facebook.

While the content of Facebook has constantly been debated, chastised, and discussed regarding what is allowed to be published, the way which Facebook moderates content has more recently come under scrutiny. In general, as the amount of content published to Facebook proliferates at an exponential rate, the company has struggled to develop a robust content moderation team or model that can scale or update itself at the same pace. As Facebook's prime goal remains with gaining more users, encouraging more content, and profiting from ad revenue, moderation seems to take a back seat, potentially viewed as a hinderance to the company wide motivation for expansion. A recent New York Times expands on an insider's look at the policies and processes in place to provide content moderation. Across multiple examples, the author found that the "billions of posts per day in over 100 languages" (Fisher) proves too overwhelming for just the fifteen-thousand-person division in charge of moderation, leading to a moderation process that is too hasty and riddled with outdated or outright mistakes. One of the sources of this problem is that the approach for setting moderation guidelines is to "try to distill highly complex issues into simple yes-or-no rules" (Fisher). As a result, when it comes to actually enforcing these policies in an environment that requires moderators to make split second decisions on content that is much richer than the policies themselves, inconsistencies and mistakes begin to bubble up across the board. In practice, many moderators have had to resort to Google Translate and are ill equipped information-wise to fully understand the scope of a post's content. The emoji slide attempts to group symbols into "indicator" groups in a very coarse-grained fashion, and often times, the results of moderation actually backfire, causing violence and discourse to accelerate in various communities. Ultimately, the lack of moderation is a worrying reveal that although Facebook encourages posts and user interaction, it has demonstrated neither the initiative nor the tools for actually understanding how the platform facilitates these interactions and monitoring it accordingly.

I believe one strategy for combating this imbalance between content and moderation is to delegate the responsibilities for setting benchmarks to a third party, preferably a public agency. My argument for this is based on a couple reasons. First, I believe that it's too difficult for social media companies to perform self-policing especially when the general view of tech companies towards moderations seems to treat it as a burden and a hassle. Juxtaposed next to common goal of generating as many users, content, ads, and revenue, and the expectation for companies to perform self-checkups becomes even more unrealistic. Put simply, the interests of a corporation cannot be trusted to align with the well-being of users and online groups, primarily because the two seem to be at odds. Instead, a third-party agency such as the government should provide a series of expectations for what can and cannot be posted online. A national baseline for content can be established, and agencies at state and city levels can enforce more localized, region specific rules. A series of rules can also be written up detailing the basics for how a company should enforce content moderation, and monthly check ups by this agency can

determine whether a company is up to standard with moderation policies, similar to how health inspectors check the hygiene standards of restaurants. Assuming the government is working in favor of the people and a company is working in favor of its own interests by creating better products and services, this dynamic between corporate and public institutions may be better suited for optimizing both the quality and safety of online forums. One of the potential barriers to entry of this approach is that if a public agency abuses its power, a government can easily enforce censorship. A case study of Chinese internet policy demonstrates how an abuse of these laws can be dangerous. I think that one way to ameliorate this issue is to write the rules clearly from the start so that nothing can be twisted to advance another party's policies. For example, the rules should be clear that "slandering" and "hate speech" cannot be argued as equivalent to "bad customer reviews" or "subpar ratings". Second, I think that because companies will try to generate and sign up as much content and users as possible, the tech companies themselves will argue against any policies that begin to censor rather than moderate. Assuming a company is not colluding with the moderation agency, the two will keep one another in check, with the company working to preserve and increase the amount of allowed content, while the content moderation agency limits lewd content, reaching a happy medium in theory.

Bibliography

Fisher, Max. "Inside Facebook's Secret Rulebook for Global Political Speech." The New York Times, The New York Times, 27 Dec. 2018, www.nytimes.com/2018/12/27/world/facebook-moderators.html.

Word Count: 891