Platforms, Speech and Truth: Policy, Policing and Impossible Choices

Blog by Mike Masnick – August 2018

So, this post was originally going to be about the choices that Facebook and other internet platforms make concerning who is allowed on their platforms, specifically in response to an interview that Mark Zuckerberg gave back in July, in which he noted that he **didn't think Facebook should remove Holocaust deniers** from its platform, saying:

> *I'm Jewish, and there's a set of people who deny that the Holocaust happened.*
>
> *I find that deeply offensive. But at the end of the day, I don't believe that our platform should take that down because I think there are things that different people get wrong. I don't think that they're intentionally getting it wrong, but I think... it's hard to impugn intent and to understand the intent. I just think, as abhorrent as some of those examples are, I think the reality is also that I get things wrong when I speak publicly. I'm sure you do. I'm sure a lot of leaders and public figures we respect do too, and I just don't think that it is the right thing to say, "We're going to take someone off the platform if they get things wrong, even multiple times."*

This created a huge furor of people talking about trolling, Holocaust denialism, Overton windows and a bunch of other things. But it's a complex, nuanced topic, and I was trying to write a complex nuanced post. And just as I was getting somewhere with it... this week, a bunch of platforms, including Apple, YouTube and Facebook, removed at least some of Alex Jones accounts or content. This created another furor in the other direction, with people talking about deplatforming, censorship, free speech, monopoly power, and policing truth. And then when Twitter chose not to follow the lead of those other platforms, we were right back to a big furor about keeping hateful wackjob conspiracy theory assholes on your platform, and whether or not you should want to do that.

Chances are no matter what I say is going to piss off pretty much everyone, but let's do the stupid thing and try to address a complex and extremely nuanced topic on the internet, with unflagging optimism that maybe (just maybe) people on the internet will (for a moment at least) hold back their kneejerk reactions of "good" or "bad" and try to think through the issues.

Let's start with a few basic principles: no matter what crazy legal analysis you may have heard before, internet sites have **every right** to remove users for nearly any reason (there may be a few limited exceptions, but none of them apply here). Whether you like it or not (and you should actually like it), corporations do get rights, and that includes their First Amendment rights to have their sites appear how they want, along with deciding who not to associate with. On top of that, again, despite what you may have heard online about Section 230 of the CDA, platforms not only have the right to moderate what's on their platform without legal liability, they are actually encouraged to do so by that law.

Indeed, if anyone knows this, it's Alex Jones, since Infowars' own terms of service **makes it clear** that Infowars can boot anyone it wants … there's a long list of rules and then it says:

> ***If you violate these rules, your posts and/or user name will be deleted. Remember: you are a guest here. It is not censorship if you violate the rules and your post is deleted. All***

> ***civilizations have rules and if you violate them you can expect to be ostracized from the tribe.***

One of the rare cases where I can say that, hey, that Alex Jones guy is absolutely right about that (and we'll leave aside the hypocrisy about him now flipping out about other sites applying those same rules on him).

A separate point that also is important, and gets regularly ignored, is that "banning" someone from these platforms often has the opposite impact of what was intended. Depending on the situation, it might not quite be a "Streisand Effect" situation, but it does create a martyr situation, which supporters will automatically use to double down on their belief that they're in the right position, and people are trying to "suppress the truth" or whatever. Also, sometimes it's **useful** to have "bad" speech out in the open, where people can track it, understand it... and maybe even counter it. Indeed, often hiding that bad speech not only lets it fester, but dulls our ability to counter it, respond to it and understand who is spreading such info (and how widely).

So, really, the question comes down to whether or not these platforms **should** be removing these kinds of accounts. But, before we should even answer that question, there's a separate question: which is **what options are there for platforms to deal with content that they disfavor?** Unfortunately, many people assume that it's a binary choice. You either keep the content up, or you take it down. But that hardly gets at the long list of possible alternatives. You can encourage good behavior and discourage bad behavior (say, with prompts if the system senses you're doing something bad, or with reminders, or by a community calling you out for bad behavior or lots of other options). Depending on the platform, you can minimize the accessibility or findability of certain content. You can minimize the reach of certain content. You can append additional information or put a "warning flag" on content. You can "shadow ban" content. You can promote "good" content to go with any content you deem to be bad. Or you can do nothing. Or you can set things up so that your users are able to help promote or minimize good or bad content. Or you can create tools that allow your users to set their own preferences and thresholds. Or you can allow third parties to build tools that do the same thing. The list goes on and on and on.

nd, yet, so much of this debate seems to ignore much of this (other than **shadowbanning**, which some people pretend is somehow evil and unfair). And, indeed, what concerns me is that while various platforms have tried some combinations of all of these things, very few seem to have really committed to these ideas -- and just get bounced back and forth between extreme pressure on two sides: "ban all the assholes" v. "how dare you fucking censor my favorite idiot."

So with the question of Alex Jones or holocaust deniers, internet platforms (again) have every right to kick them off their platforms. They don't want to be associated with assholes? Good for them. But, at the same time, it's more than a bit uncomfortable to think that anyone should want these giant internet platforms deciding who can use their platforms -- especially when having access to those platforms often feels close to necessary to take part in modern day life*. It's especially concerning when it reaches the level that basically online mobs can "demand" that someone be removed. And this is especially worrisome when many of the decisions are being made based on the claim of "hate speech," a term that not only has an amorphous and ever-changing definition, but one that has a long history of **being abused** against at risk groups or those the government simply dislikes (i.e., for those who advocate for rules against "hate speech" think about what happens when the person you trust the least gets to write the definition).

> ** Quick aside to you if you're that guy rushing down to the comments to say something like "No one needs to use Facebook. I don't use Facebook." Shut up. Most people do use Facebook. And for many people it is important to their lives. In some cases, there are necessary services that*

> *require Facebook. And you should support that rather than getting all preachy about your own life choices, good or bad.*

On top of that, I think that most people literally cannot comprehend both the scale and complexity of the decision making here when platforms are tasked with making these decisions. Figuring out which pieces of content are "okay" and which are "bad" can work when you're looking at a couple dozen piece of content. But how about a million pieces of content every single day? Or more? Back in May, when we ran a live audience "game" in which we asked everyone at a Content Moderation Summit to judge just eight examples of content to moderate, what was striking was that out of this group of **professionals in this space there was no agreement** on how to handle any piece of content. Everyone had arguments for why each piece of content should stay up, be taken down, or have flag appended to it. So, not only do you have millions of pieces of content to judge, you have a very subjective standard, and a bunch of individuals who have to make those judgment calls -- often with little training and very little time to review or to get context.

Antonio Garcia Martinez, who worked at Facebook for a while, and has been a fairly outspoken critic of his former employer (writing an entire book about it) has reasonably warned that we should be **quite careful what we wish for** when asking Facebook to cut off speech, noting that the rest of the world has struggled in every attempt to define the limits of hate speech, and it's an involved and troubling process -- and yet, many people are fine with handing that over to a group of people at a company they all seem to hate. Which... seems odd. Even more on point is an article in Fortune by CDT's Emma Llanso (who designed and co-ran much of that "game" we ran back at the content moderation summit), warning about **the lack of transparency** when platforms determine this kind of thing, rather than, say, the courts. As we've argued for years, the lack of transparency and the lack of due process is also a significant concern (though, when Mark Zuckerberg suggested an outside due process system, people **completely freaked out**, thinking he was arguing for a special Facebook court system).

In the end, I think banning people should be the "very last option" on the table. And you could say that since these platforms left him on for so long while they had their internal debates about him that that's what happened. But I don't think that's accurate. Because there were alternative solutions that they could have tried. As Issie Lapowsky at Wired pointed out in noting that this is an **unwinnable battle**, the "do nothing, do nothing, do nothing... ban!" approach is unsatisfying to everyone:

> *When Facebook and YouTube decided to take more responsibility for what does and doesn't belong on their platforms, they were never going to satisfy all sides. But their tortured deliberations over what to do with Jones left them with only two unenviable options: Leave him alone and tacitly defend his indefensible actions, or ban him from the world's most powerful platforms and turn him into the odious martyr he now is.*

Instead, we should be looking at stronger alternative ideas. Yair Rosenberg's suggestion in the Atlantic is for **counterprogramming**, which certainly is an appealing idea:

> *Truly tackling the problem of hateful misinformation online requires rejecting the false choice between leaving it alone or censoring it outright. The real solution is one that has not been entertained by either Zuckerberg or his critics: counter-programming hateful or misleading speech with better speech.*
> *How would this work in practice?*
>
> *Take the Facebook page of the "Committee for Open Debate on the Holocaust," a long-standing Holocaust-denial front. For years, the page has operated without any objection from Facebook, just as Zuckerberg acknowledged in his interview. Now, imagine if instead of taking it down,*

> *Facebook appended a prominent disclaimer atop the page: "This page promotes the denial of the Holocaust, the systematic 20th-century attempt to exterminate the Jewish people which left 6 million of them dead, alongside millions of political dissidents, LGBT people, and others the Nazis considered undesirable. To learn more about this history and not be misled by propaganda, visit these links to our partners at the United State Holocaust Museum and Israel's Yad Vashem."*
>
> *Obviously, this intervention would not deter a hardened Holocaust denier, but it would prevent the vast majority of normal readers who might stumble across the page and its innocuous name from being taken in. A page meant to promote anti-Semitism and misinformation would be turned into an educational tool against both.*

Meanwhile, Tim Lee, over at Ars Technica, **suggested another possible approach**, recognizing that Facebook (in particular) serves multiple functions. It hosts content, but it also promotes certain content via its algorithm. The hosting could be more neutral, while the algorithm is already not neutral (it's designed to promote the "best" content which is inherently a subjective decision). So, let bad content stay on the platform, but decrease its "signal" power:

> *It's helpful here to think of Facebook as being two separate products: a hosting product and a recommendation product (the Newsfeed). Facebook's basic approach is to apply different strategies for these different products.*
>
> *For hosting content, Facebook takes an inclusive approach, only taking down content that violates a set of clearly defined policies on issues like harassment and privacy.*
> *With the Newsfeed, by contrast, Facebook takes a more hands-on approach, downranking content it regards as low quality.*
>
> *This makes sense because the Newsfeed is fundamentally an editorial product. Facebook has an algorithm that decides which content people see first, using a wide variety of criteria. There's no reason why journalistic quality, as judged by Facebook, shouldn't be one of those criteria.*
>
> *Under Facebook's approach, publications with a long record of producing high-quality content can get bumped up toward the top of the news feed. Publications with a history of producing fake news can get bumped to the back of the line, where most Newsfeed users will never see it.*

Others, such as long-time free speech defender David French have suggested that platforms should ditch concepts like "hate speech" that are not in US law and simply **stick to the legal definitions" of what's allowed**:

> *The good news is that tech companies don't have to rely on vague, malleable and hotly contested definitions of hate speech to deal with conspiracy theorists like Mr. Jones. The far better option would be to prohibit libel or slander on their platforms.*
>
> *To be sure, this would tie their hands more: Unlike "hate speech," libel and slander have legal meanings. There is a long history of using libel and slander laws to protect especially private figures from false claims. It's properly more difficult to use those laws to punish allegations directed at public figures, but even then there are limits on intentionally false factual claims.*
>
> *It's a high bar. But it's a bar that respects the marketplace of ideas, avoids the politically charged battle over ever-shifting norms in language and culture and provides protection for aggrieved parties. Nor do tech companies have to wait for sometimes yearslong legal processes*

> *to work themselves out. They can use their greater degree of freedom to conduct their own investigations. Those investigations would rightly be based on concrete legal standards, not wholly subjective measures of offensiveness.*

That's certainly one way to go about it, but I actually think that would create all sorts of other problems as well. In short, determining what is and what is not defamation can often be a long, drawn out process involving lots and lots of lawyers advocating for each side. The idea that platforms could successfully "investigate" that on their own seems like a stretch. It would be fine for platforms to have a policy saying that if a court has adjudicated something to be defamatory, then they'll take it down (and, indeed, most platforms do have exactly that policy), but having them make their own determinations of what counts as defamation seems like a risky task, and what that would end up in a similar end state as where we are today with a lot of people angry at the "judgments from on high" with little transparency or right of appeal.

As for me, I still go back to the solution I've been discussing for years: we need to move to a world of **protocols instead of platforms**, in which transparency rules and (importantly) control is passed down away from the centralized service to the end users. Facebook should open itself up so that end users can decide what content they can see for themselves, rather than making all the decisions in Menlo Park. Ideally, Facebook (and others) should open up so that third party tools can provide their own experiences -- and then each person could **choose** the service or filtering setup that they want. People who want to suck in the firehose, including all the garbage, could do so. Others could choose other filters or other experiences. Move the power down to the ends of the network, which is what the internet was supposed to be good at in the first place. If the giant platforms won't do that, then people should build more open competitors that will (hell, those should be built anyway).

But, if they were to do that, it lets them get rid of this impossible to solve question of who gets to use their platforms, and moves the control and responsibility out to the end points. I expect that many users would quickly discover that the full firehose is unusable, and would seek alternatives that fit with how they wanted to use the platform. And, yes, that might mean some awful people create filter bubbles of nonsense and hatred, but average people could avoid those cesspools while at the same time those tasked with monitoring those kinds of idiots and their behavior could still do so.

I should note that this is a *different* solution than the one that Twitter's Jack Dorsey appeared to ham-fistedly suggest this week on his own platform, in which he suggested that **journalists need to do the work** of debunking idiots on Twitter. He's not wrong, but what an awful way to put it. Lots of people read it to mean "we set up the problem that makes this giant mess, and we'll leave it to journalists to come along and sort things out for free."

Instead, what I'm suggesting is that platforms have to get serious about moving real power out to the ends of their network so that anyone can set up systems for themselves -- or look to other third parties (or, even the original platforms themselves for a "default" or for a set of filter choices) for help. In the old days on Usenet there were killfiles. Email got swamped with spam, but there were a variety of anti-spam filters that you could plug-in to filter most of it out. There are ways to manage these complex situations that don't involve Jack Dorsey choosing who stays on the island and who gets removed this week.

Of course, this would require a fundamental shift in how these platforms operated -- and especially in how much control they had. But, given how they keep getting slammed on all sides for the decisions they both do and don't make, perhaps we're finally at a point where they'll consider this alternative. And, hey, if anyone at these big platforms wants some help thinking through these issues, feel free to contact us. These are the kinds of **projects we enjoy working on**, as crazy and impossible as they may feel.