

Insider Q&A: How YouTube decides what to ban

January 5 2020, by Rachel Lerman



This March 20, 2018, file photo shows the YouTube app on an iPad. YouTube's trust and safety team is tasked with deciding what the video site allows to reach its enormous audience, and what gets taken down. Team leader Matt Halprin constantly has to balance safety and civility with what the company considers a core mission: freedom of expression. (AP Photo/Patrick Semansky, File)

Matt Halprin, the global head of trust and safety for YouTube, has a tough job: He oversees the teams that decide what is allowed and what should be prohibited on YouTube.

The Google-owned site has come under fire recently for allowing videos that feature what many find offensive or violent, and for not doing enough to protect kids online. Halprin has to make [difficult decisions](#) to craft policies that keep the site as safe as YouTube wants it to be, while balancing what the company considers one of its core tenets: people's [free speech](#).

The Associated Press spoke recently with Halprin about how his team works. Questions and answers have been edited for length and clarity.

Q: How does your team operate?

A: They're separated into [policy](#) development and policy incubation. Policy development starts with the highest level of principles: We are an open platform. We do have a bias to allow freedom of expression on our platform and only remove content that we think is egregious and could cause real harm. We want to be a place where a variety of perspectives can be heard, and sometimes that even means things that people disagree with or are even offended by.

We kicked off a process a couple of years ago to essentially re-review all of our policies. We look at which policies seem to be most out of kilter with what our enforcement teams are telling us, the gray area cases or which policies are regulators talking about or the press asking about. As an example, in Q2 (June) we relaunched our hate speech policy.

Q: What does the process look like to make or change a policy?



This undated photo provided by Google shows Matt Halprin, the global head of trust and safety for YouTube. YouTube's trust and safety team is tasked with deciding what the video site allows to reach its enormous audience, and also what gets taken down. Team leader Halprin constantly has to balance safety and civility with what the company considers a core mission: freedom of expression. (Gabriela Hasbun/Google via AP)

A: The team first does the research and puts together the framework and essentially a proposal. Once it gets through me, then we bring in our cross-functional partners and people on public policy and public relations, in product, in legal. We often get sent back to the drawing board on a few issues. Then we go to an executive steering review, which is chaired by our chief product officer. Finally, the fourth and final step is the top executives. We have these meetings every single week.

As we go through this process, these guys are watching a ton of video examples.

Q: How do you think about balancing [free expression](#) with safety?

A: That is probably the toughest thing that we do. There is not a right answer. Not all of us agree. One person will think that, "Hey, we should have more civility. We shouldn't let something like this come up." And another person will say, "Yeah, but if you get rid of that uncivil comment, you lose some really valuable, you know, free expression or [political discourse](#)."

And so we have seriously huge debates about this. Sometimes we think that if we are not criticized by all sides for the policy, we've probably done something wrong. If you're only upsetting one side, then you probably haven't gotten it right.

Q: How do you ensure that things aren't slipping through the cracks when it comes to enforcement?

A: We've always had community guidelines and that's what defines our rules. We measure how much exposure occurs on content that we think goes against the line. And that's going down. For every workflow, for every policy, I get a measure of how accurate our reviewers have been regularly.

© 2020 The Associated Press. All rights reserved.

Citation: Insider Q&A: How YouTube decides what to ban (2020, January 5) retrieved 17 July 2024 from <https://techxplore.com/news/2020-01-insider-qa-youtube.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is

provided for information purposes only.