


Report: Facebook Insider Leaks 1,400 Pages of Guidelines for Steering Political Speech Towards The DNC



 Mark Zuckerberg offered a renewed defense of Facebook's handling of Russian misinformation efforts following a New York Times investigation

[AFP](#)

Yet another trove of internal Facebook documents has been leaked, this time the *New York Times* gained access to 1,400 pages of Facebook’s guidelines for global political speech.

A [recent report](#) from the *New York Times* has provided an insight into how Facebook polices global political speech. A trove of internal Facebook documents obtained by the *Times* shows how the social media platform is “a far more powerful arbiter of global speech than has been publicly recognized or acknowledged by the company itself.”

A Facebook employee reportedly leaked 1,400 pages of the internal guidelines because they “feared that the company was exercising too much power, with too little oversight — and making too many mistakes.” The documents reportedly consist of a “maze of PowerPoint slides” outlining rules for a large network of over 7,500 moderators to follow when dealing with political speech on the social media platform. These guidelines are reportedly reviewed every other Tuesday morning by several dozen Facebook employees.

Breitbart TV



The *Times* claims that these documents are filled with gaps, biases, and errors that have resulted in moderators allowing extremist speech to flourish in some countries while cracking down harshly on mainstream comments in others.

The *Times* provided an example of this problem:

Moderators were once told, for example, to remove fundraising appeals for volcano victims in Indonesia because a co-sponsor of the drive was on Facebook's internal list of banned groups. In Myanmar, a paperwork error allowed a prominent extremist group, accused of fomenting genocide, to stay on the platform for months. In India, moderators were mistakenly told to take down comments critical of religion.

These guidelines are set by young Facebook engineers and lawyers who attempt to distill complex political situations and statements into simple "yes or no" categories, then Facebook outsources actual moderation to other companies where unskilled workers spend their time attempting to enforce these ever-changing rules. Many of these moderators are often relying on tools such as Google Translate just to determine what is being said on Facebook's platform and whether it violates any rules.

Sara Su, a senior engineer on the News Feed, commented on the process stating: "It's not our place to correct people's speech, but we do want to enforce our community standards on our platform. When you're in our community, we want to make sure that we're balancing freedom of expression and safety." Monika

Bickert, Facebook's head of global policy management, stated that Facebook aimed to "prevent harm" and believed that they had been successful in that endeavor so far.

"We have billions of posts every day, we're identifying more and more potential violations using our technical systems," Bickert said. "At that scale, even if you're 99 percent accurate, you're going to have a lot of mistakes."

Navigating the actual documents seems like a huge task itself, Facebook says that they are only used as training material but employees claim that they are used as reference sheets on a daily basis. The *Times* outlines the complexity of the documents stating:

One document sets out several rules just to determine when a word like "martyr" or "jihad" indicates pro-terrorism speech. Another describes when discussion of a barred group should be forbidden. Words like "brother" or "comrade" probably cross the line. So do any of a dozen emojis.

The guidelines for identifying hate speech, a problem that has bedeviled Facebook, run to 200 jargon-filled, head-spinning pages. Moderators must sort a post into one of three "tiers" of severity. They must bear in mind lists like the six "designated dehumanizing comparisons," among them comparing Jews to rats.

Bickert discussed the issues they've faced compiling these documents saying: "There's a real tension here between wanting to have nuances to account for every situation, and wanting to have a set of policies we can enforce accurately and we can

explain cleanly.” Facebook does, however, consult with outside groups about what constitutes hate speech and what should be banned, “We’re not drawing these lines in a vacuum,” Bickert said.

The *Times* notes some of Facebook’s more extreme stances relating to “hate speech,” for example, right-wing groups such as the Proud Boys are banned but internal documents instruct moderators to allow users to praise the terrorist group known as the Taliban in certain situations:

In the United States, Facebook banned the Proud Boys, a far-right pro-Trump group. The company also blocked an [inflammatory ad](#), about a caravan of Central American migrants, that was produced by President Trump’s political team.

In June, according to internal emails reviewed by The Times, moderators were told to allow users to praise the Taliban — normally a forbidden practice — if they mentioned its decision to enter into a cease-fire. In another email, moderators were told to hunt down and remove rumors wrongly accusing an Israeli soldier of killing a Palestinian medic.

Jasmin Mujanovic, an expert on the Balkans, commented on Facebook’s moderation of speech stating: “Facebook’s role has become so hegemonic, so monopolistic, that it has become a force unto itself. No one entity, especially not a for-profit venture like Facebook, should have that kind of power to influence public debate and policy.”

Jonas Kaiser, a Harvard University expert on online extremism, said that for Facebook to become the arbiter of what constitutes extremism is “extremely problematic” as it “puts social networks in the position to make judgment calls that are traditionally the job of the courts.”

The full report from the *New York Times* [can be read here.](#)