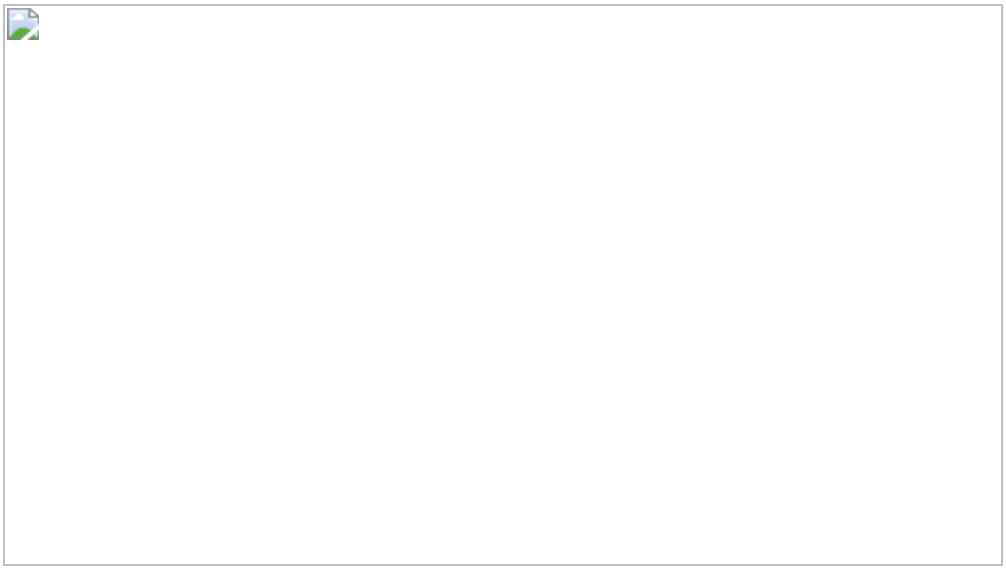
# Leaked Documents Reveal Facebooks Biased, Censorship Policies

Profile picture for user Tyler Durden
by Tyler Durden

Facebook's thousands of moderators have been relying on outdated, inaccurate and biased "maze of PowerPoint slides" to police global political speech, according to a trove of 1,400 internal documents obtained by the *New York Times*.



Moderators say they often rely on Google Translate to read posts, while facing pressure to make decisions on acceptable content within a matter of seconds, according to the report.

The guidelines - which are reportedly reviewed every other Tuesday morning by "several dozen Facebook employees who

gather over breakfast," are filled with "**numerous gaps, baises and outright errors,**" according to the *Times*.

> Moderators were once told, for example, to remove fund-raising 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 flag for possible removal comments critical of religion. -NYT

The guidelines, set by "mostly young engineers and lawyers," must be interpreted by Facebook's fleet of mostly outsourced moderators **which employ largely unskilled workers, "many hiredo out of call centers."**

> Moderators express frustration at rules they say don't always make sense and sometimes require them to leave up posts they fear could lead to violence. "You feel like you killed someone by not acting," one said, speaking on the condition of anonymity because he had signed a nondisclosure agreement. -NYT

According to Facebook executives, they are working diligently to rid the platform of "dangerous" content.

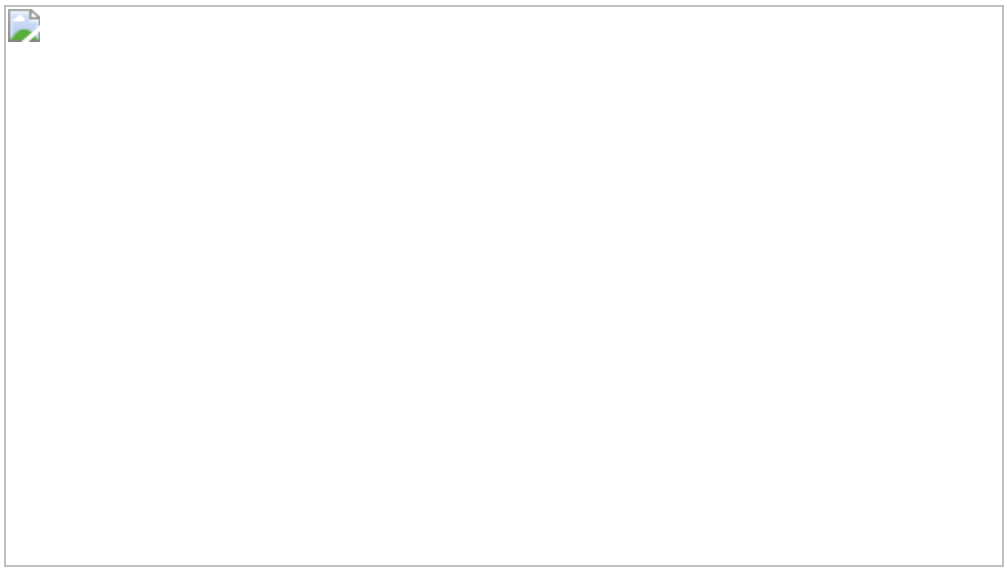
"It's not our place to correct people's speech, but we do want to enforce our community standards on our platform," said Facebook senior News Feed engineer. "When you're in our community, we want to make sure that we're balancing freedom of expression and safety."

The company's head of global policy management, Monika Bickert, meanwhile, said that the company's primary goal was to prevent harm - though perfection "is not possible."

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

And since Facebook's set of rules are more or less a patchwork of Excel spreadsheets and unorganized PowerPoint presentations, **there is no single master file** or reference guide.

> Facebook says the files are only for training, but moderators say they are used as day-to-day reference materials.
>
> Taken individually, each rule might make sense. But in their byzantine totality, they can be a bit baffling.
>
> **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**. -NYT



According to the *Times,* "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."

"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," said Bickert, who added "We’re not drawing these lines in a vacuum."

**Unseen branch of government?**

The *Times* notes that Facebook's policing of what they consider extremism or disinformation intrudes into sensitive political matters worldwide - "sometimes clumsily."

> Increasingly, the decisions on what posts should be barred amount to regulating political speech — and not just on the fringes. In many countries, extremism and the mainstream are blurring.

> 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 -NYT

"Facebook's role has become so hegemonic, so monopolistic, that it has become a force unto itself," said Balkans expert Jasmin Mujanovic. "**No one entity, especially not a for-profit venture like Facebook, should have that kind of power to influence public debate and policy.**"

During Pakistan's July elections, Facebook handed its moderators a 40-page document describing "political parties, expected trends and guidelines," which were used to **shape conversations over the primary social media platform used for news and discussions during voting**.

> The document most likely shaped those conversations — even if Pakistanis themselves had no way of knowing it. Moderators were urged, in one instance, to apply extra scrutiny to Jamiat Ulema-e-Islam, a hard-line religious party. But another religious party, Jamaat-e-Islami, was described as "benign."

> Though Facebook says its focus is protecting users, the documents suggest that other concerns come into play. Pakistan guidelines warn moderators against creating a “PR fire” by taking any action that could “have a negative impact on Facebook’s reputation or even put the company at legal risk.” -NYT

**Walking on eggshells**

Also of concern are moderator guidelines which misinterpret religious-based laws, such as one slide which tells moderators that any post degrading an entire religion is in violation of Indian law and should be flagged for removal. **This is not accurate**, according to Indian legal scholar Chinmayi Arun, who said that Indian law only prohibits blasphemy in certain conditions - such as when the speaker **intends to stoke violence**.

Another inaccurate slide says that Indian law prohibits people calling for an independent Kashmir - and moderators are specifically told to watch for posts which include the phrase "Free Kashmir," despite the common slogan among activists being completely legal.

In the Balkans, Facebook moderators were given **bad information** that Bosnian war criminal Ratko Mladic was still a fugitive, despite the fact that he was arrested in 2011.

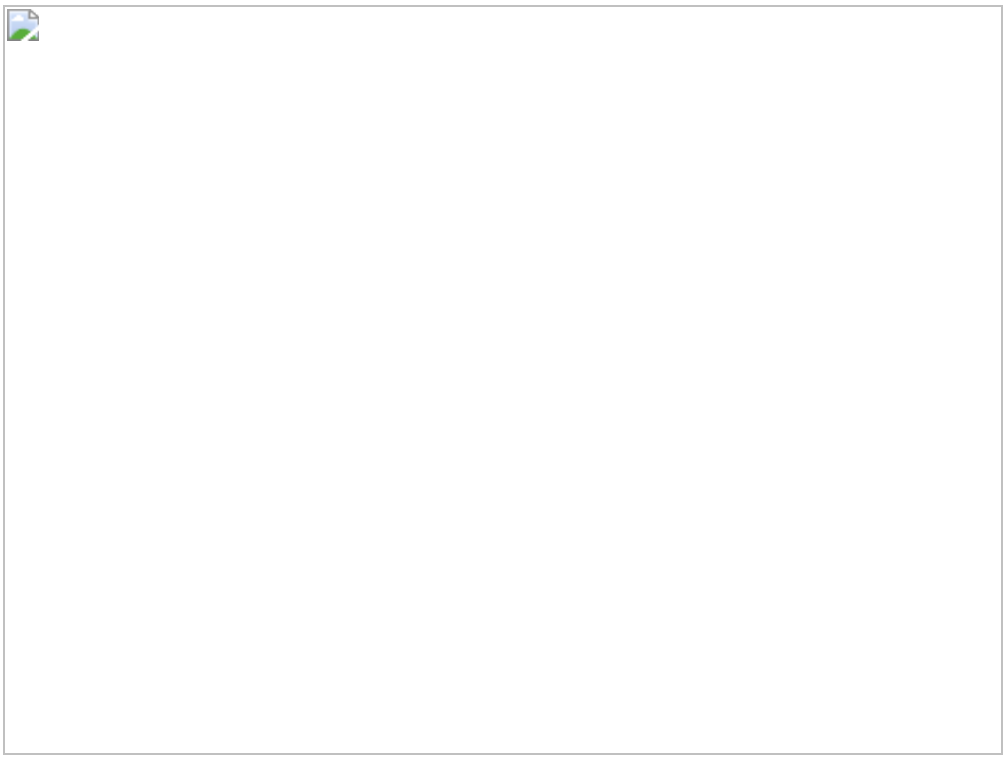
> The slides are apparently written for English speakers relying on Google Translate, suggesting that Facebook remains short on moderators who speak local languages — and who might understand local contexts crucial for identifying inflammatory speech. And Google Translate can

be unreliable: Mr. Mladic is referred to in one slide as “Rodney Young.”

**The guidelines, said Mr. Mujanovic, the Balkans expert, appear dangerously out of date**. They have little to say about ultranationalist groups stoking political violence in the region.

Nearly every Facebook employee who spoke to The Times cited, as proof of the company’s competence, its response after the United Nations accused the platform of exacerbating genocide in Myanmar. The employees pointed to Facebook’s ban this spring on any positive mention of Ma Ba Tha, an extremist group that has been using the platform to incite violence against Muslims since 2014.

But puzzled activists in Myanmar say that, months later, **posts supporting the group remain widespread.** -NYT



**The hate list**

Perhaps most politically significant of all the moderation materials is an Excel spreadsheet which includes the names of every group and individual Facebook has quietly deemed to be a hate figure. Moderators have been explicitly instructed to remove any post praising or supporting any listed figure.

> Anton Shekhovtsov, an expert in far-right groups, said he was “confused about the methodology.” The company bans an impressive array of American and British groups, he said, but relatively few in countries where the far right can be more violent, particularly Russia or Ukraine.
>
> Countries where Facebook faces government pressure seem to be better covered than those where it does not. Facebook blocks dozens of far-right groups in Germany, where the

authorities scrutinize the social network, but only one in neighboring Austria.

The list includes a growing number of groups with one foot in the political mainstream, like the far-right Golden Dawn, which holds seats in the Greek and European Union parliaments.

For a tech company to draw these lines is "**extremely problematic**," said Jonas Kaiser, a Harvard University expert on online extremism. "**It puts social networks in the position to make judgment calls that are traditionally the job of the courts.**" -NYT

The full *New York Times* report can be read here.