

This copy is for your personal, non-commercial use only. To order presentation-ready copies for distribution to your colleagues, clients or customers visit <https://www.djreprints.com>.

<https://www.wsj.com/articles/facebook-services-are-used-to-spread-religious-hatred-in-india-internal-documents-show-11635016354>

Facebook Services Are Used to Spread Religious Hatred in India, Internal Documents Show

Company researchers say anti-Muslim material is rife and calls to violence coincided with deadly riots last year. Facebook says it's working to improve enforcement.

By [Newley Purnell](#) [Follow](#) and [Jeff Horwitz](#) [Follow](#)

Oct. 23, 2021 3:12 pm ET

Mark Zuckerberg praised India in December as a special and important country for Facebook Inc., saying that millions of people there use its platforms every day to stay in touch with family and friends. Internally, researchers were painting a different picture: Facebook's products in India were awash with inflammatory content that one report linked to deadly religious riots.

Inflammatory content on Facebook spiked 300% above previous levels at times during the months following December 2019, a period in which religious protests swept India, researchers wrote in a July 2020 report that was reviewed by The Wall Street Journal.

Rumors and calls to violence spread particularly on Facebook's WhatsApp messaging service in late February 2020, when communal violence in Delhi left 53 dead, according to the report. India is Facebook's biggest market with hundreds of millions of users.

Hindu and Muslim users in India say they are subjected to "a large amount of content that encourages conflict, hatred and violence on Facebook and WhatsApp," such as material blaming Muslims for the spread of Covid-19 and assertions that Muslim men are targeting Hindu women for marriage as a "form of Muslim takeover" of the country, the researchers found.

from the files

tl;dr

| Findings | Recommendations |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>Most participants from Hindu and Muslim communities felt that they saw a large amount of content that encourages conflict, hatred and violence on Facebook and WhatsApp.</p> <p>Many Hindu and Muslim participants considered Facebook responsible for moderating this content.</p> <p>Participants shared several fear-mongering content themes: physical threat, symbolic threats, COVID-related, and false reports of communal violence among them.</p> | <ul style="list-style-type: none"> • Further develop inflammatory classifiers to detect and enforce on inflammatory content in India. • Improve our image text models so we can identify inflammatory content more effectively. • Introduce content reporting on WhatsApp and develop banks for high severity content. |
| | <ul style="list-style-type: none"> • Develop a policy that makes it possible to demote fear-mongering organic content (in-progress). • Develop country specific banks for inflammatory content and harmful misinformation relevant to At Risk Countries (ARCs). |

Source: Facebook July 2020 internal report titled, 'Communal Conflict in India Part I'

Private Facebook groups made up of like-minded users generated more divisive content. Inflammatory content primarily targeted Muslims, the researchers wrote.

Facebook was so concerned about how its services were tied to communal conflict that it dispatched researchers to interview dozens of users. A Hindu man in Delhi told them he received frequent messages on Facebook and WhatsApp “that are all very dangerous,” such as “Hindus are in danger, Muslims are about to kill us,” the researchers reported.

There is “so much hatred going on” on Facebook, one Muslim man in Mumbai was quoted as telling the researchers, saying he feared for his life. “It’s scary, it’s really scary.”

Many of the users believed it was “Facebook’s responsibility to reduce this content” in their feeds and on WhatsApp, the report said.

from the files

Key crisis events in India: 2019-20

| Event | Offline Harm | Related Content |
|---------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|
| Citizen Amendment Act (CAA) protests and violence (December 2019-February 2020) | <u>Violent episodes</u> in multiple Indian states; Police violence against protesters | Misinformation on protests, demonizing content, hate speech and inflammatory spikes |
| Delhi Riots (Feb 24-30 2020) <i>Also related to CAA protests</i> | 53 (predominantly Muslim) <u>casualties</u> ; Property damage; Police violence or inaction during riots | Rumors and calls to violence, particularly on WhatsApp |
| COVID-19 and related Anti-Muslim inflammatory content (March 2020 onwards) | <u>Violence</u> against Muslims and boycotts of Muslim businesses | <u>Green Lantern</u> : fear-mongering content blaming Muslims for the spread of Covid-19 |

Source: Facebook July 2020 internal report titled, 'Communal Conflict in India Part I'

Facebook researchers determined that two Hindu nationalist groups with ties to India's ruling political party post inflammatory anti-Muslim content on the platform, according to two separate reports earlier this year by teams investigating abuse of the company's services. The researchers recommended one of the organizations be kicked off for violating the company's hate speech rules, according to one report, but the group remains active.

The other group, researchers said, promotes incitements to violence including "dehumanizing posts comparing Muslims to 'pigs' and 'dogs' and misinformation claiming the Quran calls for men to rape their female family members." That group also remains active on Facebook, and wasn't designated as dangerous due to "political sensitivities," the report said.

The reports show that Facebook is privately aware that people in its largest market are targeted with inflammatory content, and that users say the company isn't protecting them. The documents are part of an extensive array of internal Facebook communications reviewed by the Journal that offer an unparalleled look at how its rules favor elites, its algorithms breed discord, and its services are used to incite violence and target vulnerable people.

Facebook spokesman Andy Stone declined to comment on the Hindu nationalist groups' activities on Facebook, but said that the company bans groups or individuals "after following a careful, rigorous, and multidisciplinary process."

He said some of the reports were working documents containing investigative leads for discussion, rather than complete investigations, and didn't contain individual policy recommendations.

The company has invested significantly in technology to find hate speech across languages, and globally such content on the platform has been declining, he said.

"Hate speech against marginalized groups, including Muslims, is on the rise globally," he said, adding that the company is improving enforcement and committed to updating its policies.

Facebook has a lot at stake in India, with more than 300 million Facebook users and more than 400 million people on WhatsApp. The company last year said it is

investing \$5.7 billion to expand its India operations and help boost the country's nascent digital economy.

Facebook's top executive in India reports directly to the core leadership team of Mr. Zuckerberg, Facebook's chief executive officer, in Menlo Park, Calif., underscoring the country's importance to the company. And, unlike in many developing markets, Facebook has staff on the ground in India.



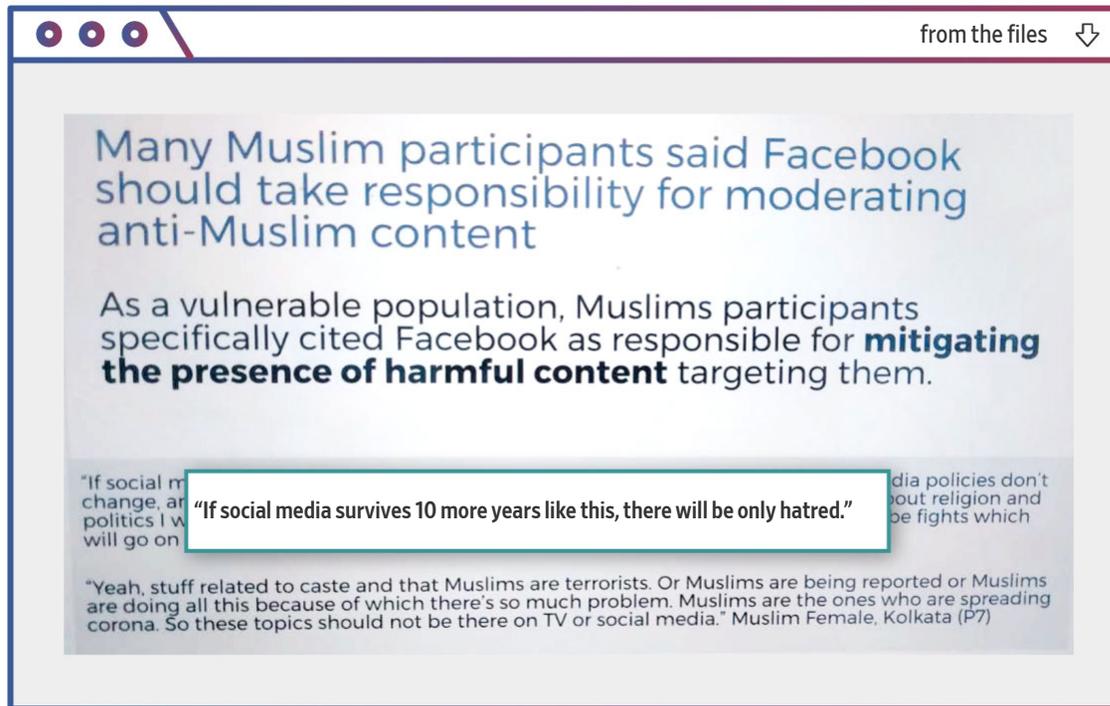
Indian Prime Minister Narendra Modi with Mark Zuckerberg at a Facebook town hall meeting at the company's Menlo Park, Calif., headquarters in 2015.

PHOTO: DAVID PAUL MORRIS/BLOOMBERG

India's government has threatened to jail employees of Facebook, WhatsApp and Twitter Inc. if they didn't comply with data or take-down requests, the Journal reported in February, as authorities sought to quash political protests and control speech on foreign-owned tech platforms. India's Ministry of Electronics and Information Technology didn't respond to requests for comment at the time, and later disputed the Journal's reporting.

The country of more than 1.3 billion has deep social and religious divisions that periodically erupt into fatal confrontations. Users in the country speak 22 major languages, making content moderation challenging, and many people have limited digital literacy.

In the July 2020 report, called “Communal Conflict in India,” one Muslim man in Mumbai was quoted as telling researchers that “if social media survives 10 more years like this, there will be only hatred.” Unless the company does a better job of policing content that encourages conflict, India will be a “very difficult place to survive for everyone,” he said, according to the report.



Source: Facebook July 2020 internal report titled, 'Communal Conflict in India Part I'

Among the researchers' recommendations for tackling inflammatory content was that Facebook invest more in resources to build out underlying technical systems that are supposed to "detect and enforce on inflammatory content in India," the way human reviewers might.

Researchers also suggested creating a "bank" of inflammatory material to study what people were posting, and creating a reporting system within Facebook's WhatsApp messaging app that would allow users to flag specific offending messages and categorize them by their contents.

Mr. Stone declined to comment on the proposals. He said Facebook invests in "internal research to proactively identify where we can improve" over the long term. Given Facebook's massive global audience, "everything that is good, bad, and ugly in our societies will find expression on our platform," but "we don't benefit from hate," he said.

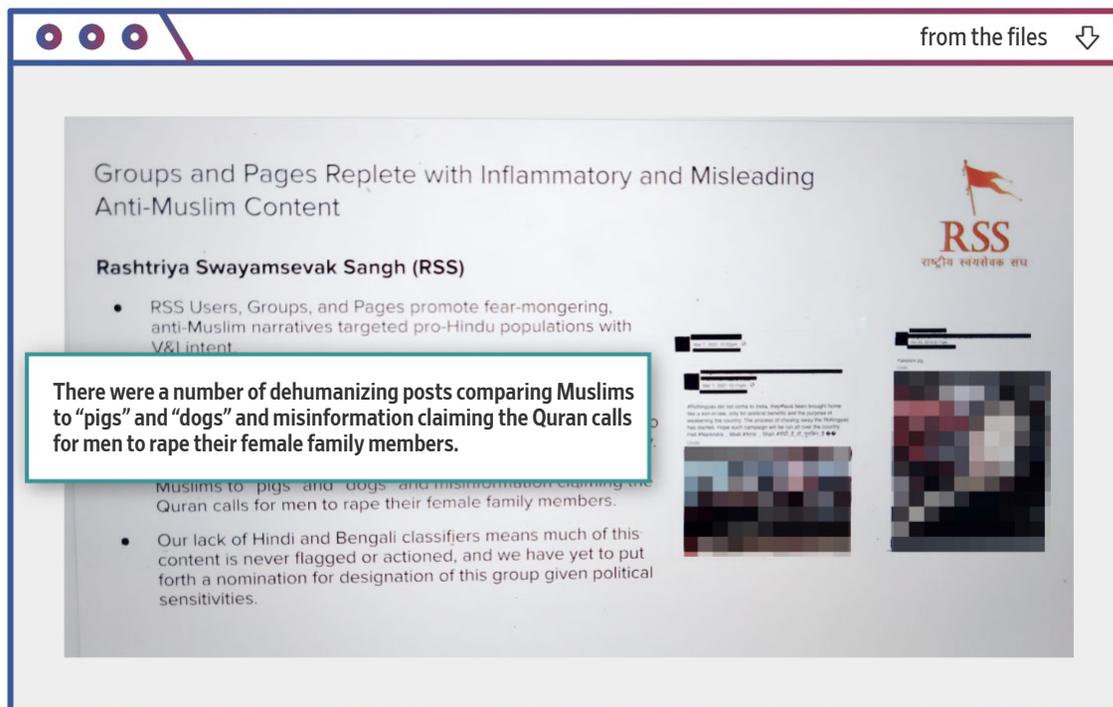
The researchers, in a report early this year called "Adversarial Harmful Networks: India Case Study," said that much of the content posted by users, groups and pages from the Hindu nationalist Rashtriya Swayamsevak Sangh group, or RSS, is never flagged. That's because Facebook lacks sufficient technical systems for detecting material in the Hindi and Bengali languages, the researchers said.

Mr. Stone said Facebook has technical systems in place to catch offending material in five languages spoken in India, including Hindi and Bengali, and has human

language expertise in many more languages. He said the company continues to work to improve its systems.

Researchers also found pro-RSS users on Facebook were posting a high volume of content about “Love Jihad,” a conspiracy theory that is spreading online and has gained traction in recent years. Proponents of the theory claim there is a plot by Muslim men to lure Hindu women with the promise of marriage or love, in order to convert these women to Islam.

Other material highlighted as inflammatory that Facebook found to be propagated by pro-RSS entities included assertions that “Muslim clerics spit on food to either ‘make it halal,’ or spread Covid-19, as a larger war against Hindus,” the researchers wrote.



Note: Third-party images have been pixelated and some names redacted. An RSS spokesman declined to comment on Facebook’s internal documents.

Source: 2021 Facebook internal report titled, ‘Adversarial Harmful Networks: India Case Study’

Facebook hasn’t designated the RSS for removal “given political sensitivities,” the document says. Prime Minister Narendra Modi worked for the RSS for decades.

A spokesman for the RSS declined to comment on Facebook’s internal documents, or the RSS’s use of Facebook’s services. “Facebook can approach the RSS anytime” about any issues related to the group’s activities, but hasn’t done so, he said.

A spokesman for Mr. Modi’s office didn’t respond to requests for comment.



Thousands of people took part in the funeral procession for two brothers who were killed during the wave of deadly violence in Delhi last year.

PHOTO: JAVED SULTAN/ANADOLU AGENCY/GETTY IMAGES

In a separate report this year, researchers said another Hindu nationalist organization called the Bajrang Dal had previously used WhatsApp to “organize and incite violence.”

One report said the group is linked to Mr. Modi’s political party, and had been considered for designation as a dangerous group, which would result in a permanent ban, and listed under a recommendation: “TAKEDOWN.” Bajrang Dal remains active on Facebook.

A spokesman for Mr. Modi’s political party, the Bharatiya Janata Party, declined to comment.

Facebook had balked at removing the Bajrang Dal from the platform following warnings in a report from its security team that cracking down on the group

might endanger both the company's business prospects and its staff in India and risk infuriating Mr. Modi's political party, the Journal reported last year.

Facebook said at the time that the company enforces its policy on dangerous individuals and organizations globally without regard to political position or party affiliation. A Bajrang Dal spokesman said at the time that its members don't take part in illegal activities and that it doesn't have conflicts with other religious groups.

Asked about the internal Facebook reports, a spokesman for the Bajrang Dal said its members don't promote hate speech on Facebook. "If they say we have broken the rules, why haven't they removed us?" he said.

The Journal reported last year that Facebook's top public-policy executive in India opposed applying Facebook's hate-speech rules to a Hindu nationalist politician, along with at least three other Hindu nationalist individuals and groups flagged internally for promoting or participating in violence, according to current and former employees. The executive, Ankhi Das, didn't respond to requests for comment at the time.

She later departed the company, saying in an internal post provided by Facebook that she had decided to step down to pursue her interest in public service. The company last month announced the appointment of a new head of public policy in India.

—Design by Andrew Levinson. A color filter has been used on some photos.

Write to Newley Purnell at newley.purnell@wsj.com and Jeff Horwitz at Jeff.Horwitz@wsj.com

Appeared in the October 25, 2021, print edition as 'Indian Users Of Facebook Fanned Hate.'

Copyright © 2022 Dow Jones & Company, Inc. All Rights Reserved

This copy is for your personal, non-commercial use only. To order presentation-ready copies for distribution to your colleagues, clients or customers visit <https://www.djreprints.com>.