

Game of Trolls

The Sinister Face of Facebook

Facebook, which was sold as a medium wherein citizens could counter government propaganda, now provides tools to governments to squash dissent.



AP Images

Facebook fail: Fewer people are eager to give Facebook thumbs up, owing to the company's policies of censoring content.

by **Tatiana Christy**

Oppressive governments around the world have realized that social media have become powerful alternative voices to official propaganda, so governments have not stopped looking for ways to combat political and social dissent on those media. Governments persecute dissenting voices either by openly denying access to the Internet and social media or by harassing critics, especially during elections or turbulent events. We have seen such trends in Africa, Asia,

Tatiana Christy is a journalist and political analyst. She is the former managing editor at the Institute for the Study of Conflict, Ideology and Policy at Boston University. She is an expert on Russia and Eastern Europe.

and the Middle East, in countries such as Russia, China, Turkey, and many others. Such governments quickly craft laws to justify their crackdown on Internet communication, wrapping them under notions of “public safety,” “terrorism,” “extremism,” “traditional values,” “treason,” “indecentcy,” etc.

However, lately oppressive governments have discovered a more subtle, but equally effective, way to censor and harass political opponents by using Facebook's Community Standards as a weapon against dissent.

When Facebook and Twitter were founded a decade ago, they heralded a new era in which the voices of ordinary citizens could be heard alongside — or even above — those of establishment insiders. From the Arab Spring and Occupy Wall

Street to demonstrations against Russia's Vladimir Putin, activists have used social media to attract followers and broadcast their messages free from official oversight. But increasingly, authoritarian regimes are deploying social media to disseminate official propaganda and crack down on dissent. What began as a tool of freedom is being turned into a weapon of repression.

In Facebook's case, Pandora's box was opened by its Community Standards, the company's rules for determining what is permissible. Facebook says it is intolerant of “nudity, hate speech, self-harm, dangerous organisations, bullying and harassment, sexual violence and exploitation, criminal activity, violence and graphic content.” And its “hate speech,” for example, covers content that directly attacks people based on their “race, ethnicity, national origin, religion, sexual orientation, sex, gender or gender identity, or serious disabilities or diseases.” If the company finds violators of its standards, it generally blocks the content.

And those Community Standards are being exploited to attack prominent voices of those who herald positive change. How it works: Repressive governments hire armies of trolls — either employed by the governments directly or by seemingly independent agencies — who go after government critics. The trolls single out Facebook accounts and, if they find any key words or phrases that Facebook might consider a “violation” of Community Standards, complain to Facebook about violations. If Facebook agrees with the complainant, one's Facebook profile can simply disappear.

Problem Platform

The core of the problem is that Facebook is unreachable to request a formal audit or to obtain information about who complained about certain posts or profiles. A Facebook profile can disappear without any possibility of users defending themselves. Some people have tried to fight the Facebook system by contacting media or NGOs dealing with free speech and human rights. Some looked for political help from the European Parliament or the European Commission, or writing directly to the very few official contacts Facebook has posted, but the bottom line is that such efforts often prove futile. The company's rules of conduct have turned into a powerful tool of censorship.

How it works: Repressive governments hire armies of trolls — either employed by the governments directly or by seemingly independent agencies — who go after government critics.

It wasn't supposed to be this way. Back in 2012, as Facebook prepared for its initial public offering, Mark Zuckerberg wrote a letter to investors touting the company's role in helping ordinary citizens hold their leaders accountable. "By giving people the power to share, we are starting to see people make their voices heard on a different scale from what has historically been possible," he wrote.

Social media has undeniably helped activist movements draw attention to their causes. But regimes around the world have figured out how to use social media to drown out dissent.

According to Freedom House's "Freedom of the Net 2017" report, fake news and trolls have led to a decline in global Internet freedom. In their yearly Freedom of the Net report, Freedom House studied 65 countries worldwide between June 2016 and May 2017 and found that Internet freedom has declined for the seventh consecutive year. In the case of Turkey, up

to 6,000 trolls (real people) have been recruited by the ruling Justice and Development Party to infiltrate online discussions in order to spread propaganda and to identify anyone criticizing the government.

"The use of paid commentators and political bots to spread government propaganda was pioneered by China and Russia but has now gone global," said Michael J. Abramowitz, president of Freedom House.

Russia does remain at the center of the discussion over the use of trolls and the dissemination of fake news. The report cites one of the main players — the Internet Research Agency, a "troll farm" with links to Russian President Vladimir Putin. The report states that the Kremlin has long used such tactics to influence domestic politics by smearing opposition figures and faking grassroots support of the government. Since 2012, Russian authorities have unjustifiably prosecuted dozens of people for criminal offenses on the basis of social media posts, online videos, media articles,

and interviews, and shut down or blocked access to hundreds of websites. Russian authorities have also pushed through parliament a raft of repressive laws regulating Internet content and infrastructure.

Russia's Internet war has spread globally, but is especially visible in its former spheres of influence in Europe, which Russia is desperate to get back. Its biggest targets are Ukraine and the former countries of the Warsaw Pact.

Some governments in the region, such as the one in Bulgaria, which is criticized by watchdog human rights and media organizations for widespread corruption and suppressing freedom of speech, have adopted authoritarian tactics on cracking down on political dissent. In Bulgaria, Facebook pages of prominent journalists and political and human rights activists have become constant targets of trolls. Accounts of such individuals are continually being reported to Facebook and blocked or erased for supposedly "violating" Facebook Community Standards. Sources say that there are massive troll factories functioning on the territory of Bulgaria, targeting government critics. This writer has requested an official statement from the Bulgarian Interior Ministry and the country's National Security Agency about these alleged troll factories, but both government entities have declined to comment. "If a government of a member of the European Union declines to respond to media questions on such essential topics as widespread censorship and suppression of free speech, it means that there is something ominous happening there," says Glenn Maxham, a former PBS anchor.

One of the many victims of troll attacks in Bulgaria is Atanas Chobanov, the editor-in-chief of the investigative reporting site Bivol. Chobanov is an award-winning journalist who writes about corruption and the abuse of power in Bulgaria. His account has been blocked multiple times by Facebook, based on reports of trolls. As a result Chobanov decided to stop using Facebook. He is one of the many examples of those abused by the platform. "I refuse to fight Facebook," he says. "It has adopted totalitarian methods in its platform and it's doing a disservice to democracy. It lacks transparency and it's unwilling to open channels for communication, where affected individuals can report troll activity and can be objectively reviewed and



Full of ... workers: A Russian "troll factory" in St. Petersburg, where hundreds of people work around the clock, infiltrating the Internet, adding praise for President Vladimir Putin and his government and smearing political opponents and the West.

AP Images

audited. In this respect Facebook has become a dangerous monster,” states Chobanov.

A similar fate has fallen upon two Lithuanian investigative reporters — Dovydas Pancerovas and Sarunas Cherniauskas — whose accounts have been blocked after trolls reported them for “hate speech.” In reality, their posts were against hate speech. “Trolls reported my posts for hate speech. Ironically, those were posts against hate speech ... and Facebook just blocked me without even looking into the post. This is dangerous,” says Cherniauskas.

Trolls dig into someone’s account looking for words or phrases or pictures that might violate community standards. Some words may be in quotes or taken out of context; some are used to convey an opposite meaning — a message of tolerance — but still Facebook’s system finds them to be a “violation.” The most common blocks occur when someone is reported for using the word “gay” or its derivatives. Trolls report posts that use this word in the context of “hate speech.”

Facebook and Fallacies

A Facebook spokesman, Christine Chen, said that anyone can report material that violates Facebook Community Standards, but that the complaints are verified by Facebook. “When we look at something that has been flagged, it does not matter how many reports we receive about an individual post or profile, nor where the report comes from. Our teams review reports based on our policies, and will remove the material if there is a violation.”

However, according to Facebook sources, when a Facebook user is the subject of a great number of reports, this may signal reviewers that something needs to be prioritized for a more accelerated process, often resulting in the removal of content. In other words, the number of reports also matters. And this comes in really handy to trolls, who can massively batter a single profile in minutes through organized reporting.



We censor because we care: Facebook’s commitment to remove hate speech from its platform comes with armies of trolls, unscrupulously using Facebook’s Community Standards to stifle dissent.

There is a myth that Facebook accounts are blocked automatically by algorithms, without human intervention. On the contrary, there are real people who are part of the Facebook Community Operations team looking at reported content. These people span the globe and speak more than 40 languages.

So far, the company has been very unwilling to disclose details about the way content reviewers are selected and monitored. There are reports that Facebook outsources this activity to third parties, who then hire content moderators — such as one entity in Germany described by Max Hoppenstedt in his article “A Visit to Facebook’s Recently Opened Center for Deleting Content.”

The clandestine practices surrounding the operations of these centers are a cause for much speculation, leading to suspicions that certain reviewers might have received “special instructions” on how to treat certain accounts or be special government recruits. The lack of transparency for where such sites operate geographically or details on how reviewers are hired has brought Facebook lots of criticism. The company, however, is unwilling to discuss

details in the hiring process of content moderators.

Facebook, for its part, denies any bad intent. “Unfortunately, we do sometimes make mistakes. Posts and accounts can sometimes be removed in error, and we restore them as soon as we are able to investigate. Our team processes millions of reports each week, and sometimes we do get things wrong. We’re very sorry about these mistakes and we welcome feedback and suggestions for improvement,” says Chen.

Unfortunately, these words lack conviction: The victims really have no place to complain. Dimitar Bojantchev, a Silicon Valley IT professional, says, “There is no effective way to communicate with Facebook. You simply don’t know where to go to complain. There is no information about this posted anywhere. The standard forms the platform offers for messaging with concerns or complaints

regarding blocked accounts do not work. Once someone is blocked, you can report ‘that this is a mistake’ a hundred times and still won’t get an answer.” He also says that, from a technical standpoint, it should not be a problem for Facebook to know who is reporting supposed Facebook violators. It should be easy to find out whether complainants are real people with real accounts or whether they are acting in concert to target specific accounts. Another clue should tip off Facebook: the age of the Facebook posts that are being complained about. Trolls dig deep into targeted profiles, looking for words that Facebook doesn’t like. Many victims have been blocked for very old posts — from several years earlier — and this should be a red flag for reviewers. But apparently it is not.

Facebook claims it does regular audits on the accuracy of reviewers’ decisions and that it periodically audits all auditors. However, there is something very broken in its system if it allows itself to be compromised so easily by attacks from the outside.

Social media are the new frontier of news dissemination, and we (as their users) need to protect them against censorship. ■