

Sharing on Facebook seems harmless, but leaked documents show how it may help spread misinformation

December 28 2021, by Rachel Axon



Credit: Unsplash/CC0 Public Domain

Dec. 28—A video of House Speaker Nancy Pelosi seeming to slur her speech at an event tore through the internet, gaining steam on Facebook.

Share after share, it spread to the point of going viral.

The altered video from May 2019 was a slowed-down version of the actual speech the California Democrat gave but was being promoted as real. Even though Facebook acknowledged the video was fake, the company allowed it to stay on the platform, where it continued to be reshared. That exponential resharing was like rocket fuel to the manipulated video.

In the run-up to the 2020 election, with additional traction coming from then-President Donald Trump sharing the video, the amplification of misinformation showed the real-world implications and the need for [social media companies](#) to take action to stem its spread.

YouTube, where it also appeared, took the video down. But Facebook said at the time because the company wanted to encourage free expression, it allowed it to remain up while reducing distribution of it to strike a balance between that priority and promoting authentic content.

The fake Pelosi video is an example of the power of something [social media users](#) do naturally—sharing.

It turns out, internal documents show, that a company researcher found that Facebook could have flagged the source of that video, the Facebook page of Politics WatchDog, at least week earlier based on a simple metric—how much traffic was coming from people sharing its content.

With its content surfacing almost exclusively from Facebook users resharing its posts, the page had gained a massive audience in the days leading up to the Pelosi video through a strategy one researcher dubbed "manufactured virality," or when a group uses content that has already gone viral elsewhere to drive their Facebook page's popularity.

While not the exclusive domain of shady intent, the approach is common by bad actors on Facebook often to spread falsehoods. Facebook has allowed this type of content to flourish on its platform.

Sharing in Facebook isn't inherently bad. It is, after all, a basic function of how social media works and why many of us go there.

What Facebook's internal research shows about sharing

In documents released by whistleblower Frances Haugen, Facebook employees warn repeatedly of the likelihood that reshares like these were a main vector for spreading misinformation and the harms that could come from that. They suggested myriad solutions—everything from demoting them to slowing them down—only to see their suggestions ignored.

Over the red flags raised by some employees, Facebook made sharing easier during that time, choosing core engagement metrics critical to its business over measures that could have reduced the harmful content on the platform. Getting people to read, share and respond to Facebook content and spend more time on the platform is critical to what the company can charge advertisers, and it found misinformation in reshares to be particularly engaging.

In a whistleblower complaint Haugen filed with the Securities and Exchange Commission, she included reshares as one of the ways Facebook has failed to remove misinformation from the platform even as it touted its efforts to do so.

While Facebook had publicized its efforts countering extremism and misinformation related to the 2020 U.S. elections and the Jan. 6

insurrection, it failed to adequately account for its role in the spread of misinformation, Haugen's complaint states.

"In reality, Facebook knew its algorithms and platforms promoted this type of harmful content," her complaint says, "and it failed to deploy internally-recommended or lasting counter-measures."

Attorneys for Haugen, a former Facebook product manager, disclosed more than 1,000 documents to the SEC and provided them to Congress in redacted form. USA TODAY was among a consortium of news organizations that received redacted versions.

The documents have shed light on internal research showing Facebook's knowledge of a variety of harms, many of which were first reported by The Wall Street Journal.

Meta Platforms, Facebook's parent company, declined to answer a list of detailed questions about misinformation spread through reshares, the solutions offered by its employees and the company's incentives not to act on reshares because of the impact on its engagement metrics.

"Our goal with features like sharing and resharing is to help people and communities stay connected with each other," Aaron Simpson, a spokesman for Meta, wrote in an emailed statement. "As with all our features and products, we have systems in place to keep communities safe, like reducing the spread of potentially harmful content."

Why sharing on Facebook can be connected to misinformation

To be sure, sharing is not inherently bad and, indeed, is a bedrock of the platform. Users do it all the time to share news of a friend facing a

medical issue, seek help finding a lost pet, announce a birth or just pass on something they found interesting.

But Facebook's research found misinformation in particular draws user engagement with a high likelihood of being reshared and that the company could use reshare indicators to lessen the reach of harmful content.

Experts agreed the key role of reshares in spreading misinformation and Facebook's inaction have not been widely known. The documents show its reluctance to reduce the spread of misinformation in reshares as doing so impacts the kind of engagement that Facebook profits from.

"One thing that we have seen consistently, not just in these documents but in other reports about actions that Facebook has taken, is that Facebook is not willing to sacrifice its business goals to improve the quality of content on its system and achieve integrity goals," said Laura Edelson, co-director of Cybersecurity for Democracy at New York University.

Facebook disabled Edelson's account after her research team created a browser extension that allows users to share information about which ads the site shows them. Other experts agreed with her assessment of Facebook's incentives playing a role in its decisions about how, and whether, to address this type of misinformation on the platform.

Edelson added, "We do see Facebook is consistently willing to sacrifice its integrity goals for the sake of its overall business goals."

The role of Facebook's algorithm as accelerant

In a late 2018 note, Meta Platforms CEO Mark Zuckerberg explained Facebook's efforts to combat misinformation, namely content that

borders on violating its policies. The closer a piece of content gets to that line, the more people engaged with it even as they said they didn't like it, he wrote.

Zuckerberg said the company would work to reduce the distribution and virality of this type of content, specifically misinformation.

Yet over and over in the documents, Facebook's employees reiterate the likelihood that reshared content is misinformation and found that these shares are a key indicator it can use to reduce the distribution of likely harmful content.

How many layers of resharing, or its reshare depth, can also be an indicator of its potential for harm. Facebook has a metric for what it calls "deep reshares."

When you post a link or a video, for instance, according to Facebook's measure, that originating post has a reshare depth of zero. Then one of your friends clicks the button to share your post, and that bumps it to a depth of one. If their friend or follower shares that, the depth is two. And so on, and so on.

Facebook found a reshare depth of two or greater for a link or photo indicated that piece of content was four times as likely to be misinformation compared to other links and photos in the news feed generally. That could increase to up to 10 times as likely to be misinformation at higher reshare depths.

That doesn't mean everything reshared six steps from the original poster is misinformation. But, according to Facebook's research, it is far more likely to be.

In a 2020 analysis, Facebook found group reshares are up to twice as

likely to be flagged as problematic or potentially problematic. Another analysis that year found that since 2018 content shared by groups grew three times faster than content shared outside of groups overall.

According to one document, up to 70% of misinformation viewership comes from people sharing what others have shared.

"If we are talking about stuff that is misinformation or hate speech that (Facebook says) they do not want to tolerate on their platform and then they just let it run wild, I'd say yes there is also something that they could and should do about it," said Matthias Spielkamp, executive director of Algorithm Watch, a research and advocacy organization.

Facebook's algorithm, optimized for engagement and virality, serves as an accelerant and further amplifies content that is gaining momentum on its own.

While individual users can create misinformation that gets reshared, Facebook's research focused on the particular harm of groups and pages—including those that use the company's algorithms as a way to spread this type of content and grow their following.

"These kind of actors who are trying to grow their celebrity status, to grow their follower networks, they understand that you make sensational content, you make stuff that really surprises people, captures their attention and trades on their already held beliefs and you keep working on that and pretty soon you've got a nice follower base," said Jennifer Stromer-Galley, a Syracuse University professor who studies social media.

Facebook's documents warn of the harms that could come from reshared misinformation. One 2019 experiment found adding friction to sharing in India reduced "particularly concerning" content that inflamed tensions

about Kashmir.

Another document from 2019 warned that "political operatives and publishers tell us that they rely more on negativity and sensationalism for distribution due to recent algorithmic changes that favor reshares."

Citing those concerns political and news actors in the United States and Europe, one document from 2020 noted that Facebook's data showed misinformation, violent content and toxicity were "inordinately prevalent among reshares."

The altered Pelosi video was exactly the type of content Facebook's algorithm incentivized, and using reshares of earlier content as a signal the company could have flagged Politics WatchDog at least a week before the video posted.

A small group of Facebook pages can have big influence

A researcher explained that through manufactured virality, a small cohort of pages commanded an outsized influence on Facebook. According to the document, half of all impressions through reshares across Facebook went to pages that got at least 75% of their impressions from reshares. Nearly a quarter of those impressions went to pages with rates of 95% or higher.

A Facebook researcher recommended flagging pages that get more than half their impressions through reshares, overriding the algorithm's automated amplifying effect and instead demoting them until manufactured virality is no longer an effective growth strategy. Facebook should instead reward original creators who work harder to earn their audiences, the researcher suggested.

It is unclear if Facebook has adopted the recommendation. The company did not answer a question about what steps it has taken to address manufactured virality.

A former Facebook employee did raise concerns about tamping down viral content.

Alec Muffett, a software engineer, left Facebook in 2016 over concerns of the company's potential expansion to China and proposals for the country's authoritarian government to be able to downrank content in feeds.

"Everybody is talking about 'harms,' but nobody is valuing the 'benefits' of free viral expression," Muffett wrote in an email. "Viral speech is a powerful phenomenon, and it constitutes the online form of 'freedom of assembly.' People are learning to adapt to modern forms of it. I am deeply concerned at any proposal that virality should be throttled or intermediated by authorities, or by platforms on behalf of authorities."

'Facebook sells attention': Could the solution be bad for business?

Facebook's deliberations of how to handle misinformation spreading through reshares inevitably circle back to one concern in the documents: They generate likes, comments and shares—exactly the kind of engagement the company wants. That incentivizes bad actors, but, to Facebook, it's also good for business.

"The dramatic increase in reshares over the past year is in large part due to our own product interventions," one document from early 2020 found.

"Reshares have been our knight in shining armor," another document

noted.

It is not in Facebook's interest to tamp down on this information, experts argued.

"It clearly says that they put their business interests over having a civilized platform" said Spielkamp, of Algorithm Watch.

"It's hard to come up with a different explanation than to say, 'We know it's gross what people are sharing and we know how we could slow it down, but we are not doing it.'"

In 2018, Facebook shifted to a key metric called meaningful social interactions (MSI). Ostensibly, the goal was to show users more content from friends and family to promote those interactions. But in doing so, it valued engagement—likes, comments and shares—and Facebook's documents found misinformation and content that generates outrage is more likely to do that.

One early explanation of meaningful social interactions among the Facebook Papers shows reshared content being weighted 15 times that of a like.

"If they're over-weighting reshares—and we know absolutely it's the case that information that is incorrect or sensational spreads at a much faster rate than correct, factual information—taking the gas out of those messages would be tremendously helpful," said Stromer-Galley.

"When the algorithm then gives that a speed boost—which is what's happening now—then that is something the tech company is responsible," said Stromer-Galley. "If they dial it back or even stop the spread completely, it's not really even that they're regulating the content....If it just happens to have a particular shape to it, then it gets

throttled."

Facebook ran an experiment in 2019, trying to reduce the spread of reshares more than two shares away from the original poster. It found lessening the spread of that content produced "significant wins" in reducing misinformation, nudity and pornography, violence, bullying and disturbing content.

That experiment found no impact on the number of daily users on Facebook, the time they spent on the platform or how many times they logged on. But it cautioned that keeping the wins on reducing negative content might require Facebook to change its goals on meaningful social interactions.

Because changes to distribution of reshares were likely to affect the company's top-line metrics, they were often escalated to leadership and involved red tape to weigh integrity improvements against engagement, one former employee said. That person agreed to speak on the condition of anonymity.

In April 2020, a Facebook team offered a presentation Zuckerberg of soft actions it could take, effectively reducing the spread of this kind of harmful content without actually taking it down. One such action proposed changes to Facebook's algorithm that had ranked content on the likelihood that people steps removed from the original poster would react, comment or share it.

Facebook was already doing this for some content, the document says, and expected a reduction of 15% to 38% of misinformation on health and civic content, which Facebook uses to describe political and social issues.

"Mark doesn't think we could go broad, but is open to testing, especially

in (at-risk countries)," a Facebook employee wrote. "We wouldn't launch if there was a material tradeoff with MSI impact."

Simpson, the Meta spokesperson, said Facebook adjusts the weight of rankings signals such as reshares "when we find a relationship with integrity concerns" and on certain topics, such as health, political or social issues.

Experts argued Facebook could take further steps to demote viral shares, but it's the structure of the platform that enables them to go viral while the company profits from that engagement. The company's documents seem to back that up.

In one document, a Facebook employee wrote, "We also have compelling evidence that our core product mechanics, such as virality, recommendations, and optimizing for engagement, are a significant part of why these types of speech flourish on the platform."

What Facebook tried to slow the spread of misinformation

Over the years, Facebook's employees have proposed several possible solutions.

One suggested demoting reshared content where the person posting it isn't connected to the original poster. That document estimated that would reduce link misinformation by a quarter and photo misinformation by half on political and social issues.

An experiment abroad showed the promise of adding obstacles to resharing. Facebook removed the share button and the whole section with reactions and comments to a post and found it reduced subsequent

viewership for misinformation by 34% and graphic violence by 35%.

Other social media platforms have been employing some efforts to stem or at least slow the spread of misinformation. Twitter, for instance, added "misleading information" warnings, restrictions on retweets with misleading information and other features adding a layer of intent—and perhaps consideration—before users could reshare content.

"I do not see Facebook prioritizing its role as an information purveyor in our democracy," said Stromer-Galley. "I don't see them taking that role seriously because if they did, then we should have seen some of these interventions actually used."

What role Facebook plays—platform, publisher, utility or something else—is a hotly debated topic, even by the company itself.

Still, Facebook did, in some instances, roll out changes—at least for a time. It demoted deep reshares in at least six countries, according to the documents.

Despite cutting the spread of photo-based misinformation by nearly 50% in Myanmar when it slowed distribution based on how far from the originator the resharing was, Facebook said it planned to "roll back this intervention" after the country's election.

Rather than widely implementing measures to limit the reach of reshares, ultimately Facebook made it easier for reshares to spread [misinformation](#) on the platform.

"There have been large efforts over the past two years to make resharing content as frictionless as possible," one document noted.

In 2019, Facebook rolled out the group multi-picker—a tool that would

allow users to share content into multiple groups at the same time. That increased group reshares 48% on iOS and 40% on Android.

As it turns out, Facebook found those reshares to be more problematic than original group posts, with 63% more negative interactions per impression. Simpson said the group multi-picker has been inactive since February.

But tools like that are ripe for abuse, experts argued.

"Facebook sells attention. Things go viral because they capture a lot of attention," Edelson said. "What the researchers are really struggling with is that the thing that is at the center of Facebook's business model is also the thing that is causing the most harm to Facebook users."

©2021 USA Today

Distributed by Tribune Content Agency, LLC.

Citation: Sharing on Facebook seems harmless, but leaked documents show how it may help spread misinformation (2021, December 28) retrieved 7 May 2024 from <https://techxplore.com/news/2021-12-facebook-harmless-leaked-documents-misinformation.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.