

EXPLAINER: How Elon Musk is changing what you see on Twitter

December 14 2022, by MATT O'BRIEN, BARBARA ORTUTAY and DAVID KLEPPER



SpaceX's Elon Musk waves while providing an update on Starship, on Feb. 10, 2022, near Brownsville, Texas. Musk's Twitter has dissolved its Trust and Safety Council, the advisory group of around 100 independent civil, human rights and other organizations that the company formed in 2016 to address hate speech, child exploitation, suicide, self-harm and other problems on the platform. The council had been scheduled to meet with Twitter representatives Monday night, Dec. 12. But Twitter informed the group via email that it was disbanding it

shortly before the meeting was to take place, according to multiple members.
Credit: Miguel Roberts/The Brownsville Herald via AP, File

What you're seeing in your feed on Twitter is changing. But how?

The social media platform's new owner, Elon Musk, has been trying to prove through giving selected journalists access to some of the company's internal communications dubbed "The Twitter Files" that officials from the previous leadership team allegedly suppressed right-wing voices.

This week, Musk disbanded a key advisory group, the Trust and Safety Council, made up of dozens of independent civil, human rights and other organizations. The company formed the council in 2016 to address hate speech, harassment, child exploitation, suicide, self-harm and other problems on the platform.

What do the developments mean for what shows up in your feed every day? For one, the moves show that Musk is prioritizing improving Twitter's perception on the U.S. political right. He's not promising unfettered free speech as much as a shift in what messages get amplified or hidden.

WHAT ARE THE TWITTER FILES?

Musk bought Twitter for \$44 billion in late October and since then has partnered with a group of handpicked journalists including former Rolling Stone writer Matt Taibbi and opinion columnist Bari Weiss. Earlier this month, they began publishing—in the form of a series of tweets—about actions that Twitter previously took against accounts thought to have violated its content rules. They've included screenshots

of emails and messaging board comments reflecting internal conversations within Twitter about those decisions.

Weiss wrote on Dec. 8 that the "authors have broad and expanding access to Twitter's files. The only condition we agreed to was that the material would first be published on Twitter."

Weiss published the fifth and most recent installment Monday about the conversations leading up to Twitter's Jan. 8, 2021 decision to permanently suspend then-President Donald Trump's account "due to the risk of further incitement of violence" following the deadly U.S. Capitol insurrection two days earlier. The internal communications show at least one unnamed staffer doubting that one of the tweets was an incitement of violence; it also reveals executives' reaction to an advocacy campaign from some employees pushing for tougher action on Trump.

WHAT'S MISSING?

Musk's Twitter Files reveal some of the internal decision-making process affecting mostly right-wing Twitter accounts that the company decided broke its rules against hateful conduct, as well as those that violated the platform's rules against spreading harmful misinformation about COVID-19.

But the reports are largely based on anecdotes about a handful of high-profile accounts and the tweets don't reveal numbers about the scale of suspensions and which views were more likely to be affected. The journalists appear to have unfettered access to the company's Slack messaging board—visible to all employees—but have relied on Twitter executives to deliver other documents.

THE TWITTER FILES MENTION SHADOWBANNING. WHAT'S THAT?

In 2018, after then-CEO Jack Dorsey said Twitter would focus on the "health" of conversations on the platform, the company outlined a new approach intended to reduce the impact of disruptive users, or trolls, by reading "behavioral signals" that tend to indicate when users are more interested in blowing up conversations than in contributing.

Twitter has long said it used a technique described internally as "visibility filtering" to reduce the reach of some accounts that might violate its rules but don't rise to the level of being suspended. But it rejected allegations it was secretly "shadowbanning" conservative viewpoints.

Screenshots showing an employee's view of prominent user accounts disclosed through the Twitter Files show how that filtering works in practice. It's also led Musk to call for changes to make that more transparent.

"Twitter is working on a software update that will show your true account status, so you know clearly if you've been shadowbanned, the reason why and how to appeal," he tweeted.

WHO'S MONITORING POSTS ON TWITTER NOW?

Musk laid off about half of Twitter's staff after he bought the platform and later eliminated an unknown number of contract workers who had focused on content moderation. Some workers who were kept on soon quit, including Yoel Roth, Twitter's former head of trust and safety.

The departure of so many employees raised questions about how the platform could enforce its policies against harmful misinformation, hate speech and threats of violence, both within the U.S. and across the globe. Automated tools can help detect spam and some suspicious accounts, but others take more careful human review.

It's likely the reductions will force Twitter to concentrate content moderation efforts on regions with stronger regulations governing [social media platforms](#) like Europe, where tech companies could face big fines under the new Digital Services Act if they don't make an effort to combat misinformation and hate speech, according to Bhaskar Chakravorti, dean of global business at the Fletcher School at Tufts University.

"The staff has been decimated," Chakravorti said. "The few content moderators left are going to be focused on Europe, because Europe is the squeakiest wheel."

HAS THERE BEEN AN IMPACT?

Since Musk bought Twitter a number of researchers and advocacy groups have pointed to an increase in posts containing racial epithets or attacks on Jewish people, gays, lesbians and [transgender people](#).

In many cases, the posts were written by users who said they were trying to [test Twitter's new boundaries](#).

According to Musk, Twitter acted quickly to reduce the overall visibility of the posts, and that overall engagement with [hate speech](#) is down since he purchased the company, a finding [disputed by researchers](#).

The most obvious sign of change at Twitter are the [formerly banned users](#) whose accounts have been reinstated, a list that includes Trump, satire site The Babylon Bee, the comedian Kathy Griffin, Canadian psychologist Jordan Peterson and, before he was kicked off again, Ye. Twitter has also reinstated accounts of neo-Nazis [white supremacists](#) including Andrew Anglin, the creator of the white supremacist website Daily Stormer—along with QAnon supporters whom Twitter's old guard had been removing in masses to prevent hate and misinformation from

spreading on the platform.

In addition, some high-profile tweeters like Republican Rep. Marjorie Taylor Greene who were previously banned for spreading misinformation about COVID-19 have resumed posting misleading claims about vaccine safety and sham cures.

Musk, [who has spread false claims about COVID-19 himself](#), returned to the topic this with a tweet this week that mocked transgender pronouns while calling for [criminal charges](#) against Dr. Anthony Fauci, the nation's top infectious disease expert and one of the leaders of the country's COVID response.

Calling himself a "[free-speech](#) absolutist," Musk has said he wants to allow all content that's legally permissible on Twitter but also that he wants to downgrade negative and hateful posts. Instead of removing toxic content, Musk's call for "freedom of speech, not freedom of reach" suggests Twitter may leave such content up without recommending it or amplifying it to other users.

But after cutting out most of Twitter's policy-making executives and outside advisers, Musk often appears to be the arbiter of what crosses the line. Last month, Musk himself announced that he was booting Ye after the rapper formerly known as Kanye West posted an image of a swastika merged with a Star of David, a post that was not illegal but deeply offensive. The move led to questions about what rules govern what can and can't be posted on the platform.

© 2022 The Associated Press. All rights reserved. This material may not be published, broadcast, rewritten or redistributed without permission.

Citation: EXPLAINER: How Elon Musk is changing what you see on Twitter (2022, December 14) retrieved 17 April 2024 from <https://techxplore.com/news/2022-12-elon-musk-twitter.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.