

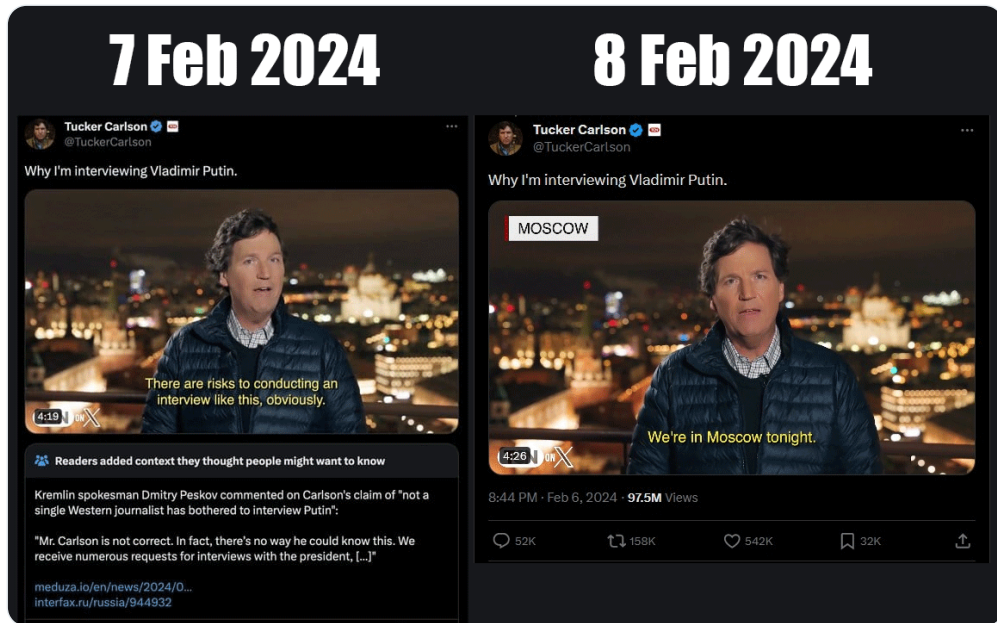


Pekka Kallioniemi @P_Kallioniemi

Feb 8, 2024 · 16 tweets · [P_Kallioniemi/status/1755566046541492630](https://twitter.com/P_Kallioniemi/status/1755566046541492630)

In today's #vatniksoup, I'll talk briefly about the Community Notes system and why it doesn't work. I've previously stated that the Community Notes mechanism is a "mob rule" and can be played easily by big accounts and troll farms.

1/15



Community Notes is a community-driven content moderation program, intended to provide informative context based on a crowd-sourced voting system. As of Nov 2023, this system had over 130 000 contributors.

2/15

Elon Musk Embraces Twitter's Radical Fact-Checking Experiment

The project allows users to suggest short notes that add missing context to viral tweets. It could change how social platforms operate.



PHOTOGRAPH: MIRAGEC/GETTY IMAGES

The idea of a crowd-sourced system as a moderation tool did not come from Elon - it was announced already back in 2020 when it was called Birdwatch. Musk later rebranded the system as Community Notes and sold it to the platform as something new.

3/15

TECH / TWITTER - X / WEB

Twitter's 'Birdwatch' looks like a new attempt to root out propaganda and misinformation



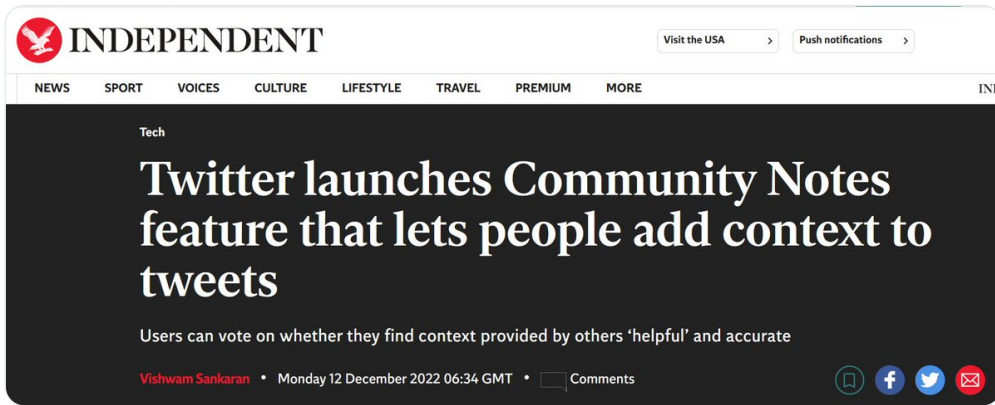
Illustration by Alex Castro / The Verge

/ Twitter confirms it will share more "soon"

By Sean Hollister, a senior editor and founding member of The Verge who covers gadgets, games, and toys. He spent 15 years editing the likes of CNET, Gizmodo, and Engadget.

Oct 4, 2020, 8:00 PM GMT-3

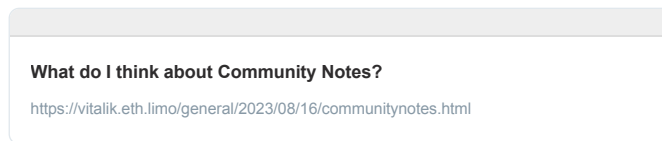




Vitalik Buterin (@VitalikButerin) has made a very extensive (and technical) analysis on the tool and the Community Notes algorithm as a whole.

I disagree with him on some points, but I really suggest to everyone to read it:

4/15



Also, focusing the analysis only on the algorithm and the technical aspects is simplifying the concept, as it rules out the most important variable: human factor.

People are prone to bias and disinformation tends to spread much more aggressively than the truth.

5/15



By
Larry
Greenemeier,
Scientific
American

Leave a
comment

False news travels 6 times faster on Twitter than truthful news

Science Mar 9, 2018 9:51 AM EST

Share

Twitter's former head of safety, Yoel Roth, has stated that the system was never intended to replace the curation team, but to complement it. But all this of course changed after Elon sacked everyone from Twitter's Trust and Safety in order to save money.

PREMIUM • EDITORS' PICK

Musk's X Fired 80% Of Engineers Working On Trust And Safety, Australian Government Says

Half of internal moderators and a third of X's trust and safety team globally were fired after Musk's arrival, Australia's online safety regulator says.

Thomas Brewster Forbes Staff

Senior writer at Forbes covering cybercrime, privacy and surveillance.

Follow

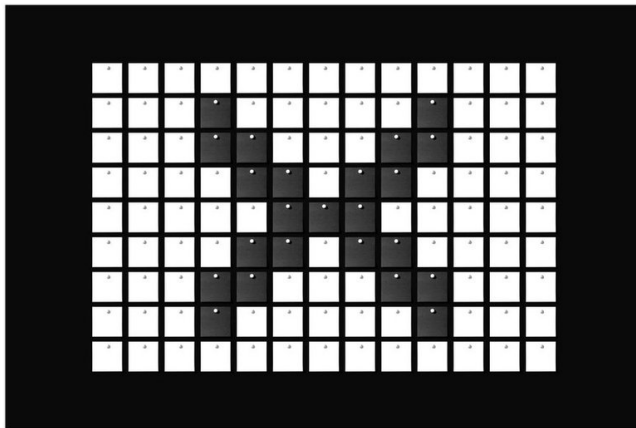


Jan 10, 2024, 08:00am EST

VICTORIA ELLIOTT DAVID SILBERG SECURITY OCT 17, 2023 8:28 PM

Elon Musk's Main Tool for Fighting Disinformation on X Is Making the Problem Worse, Insiders Claim

X is promoting Community Notes to solve its disinformation problems, but some former employees and people who currently contribute notes say it's not fit for that purpose.



PHOTOGRAPH: MIRAGEC/GETTY IMAGES

These sackings have resulted in long response times on reports on hate speech - X's attempts to deal with hateful direct messages has slowed down by 70%.

As of today, the company doesn't have any full-time staff singularly dedicated to hateful conduct issues globally.

7/15

The revelations come as the commissioner used Australia's Online Safety Act to demand X answer questions on how it is keeping users safe. Musk's company said that prior to his \$40 billion acquisition in October 2022, it had 279 engineers focused on trust and safety globally. That dropped to 55 by the end of May 2023. Its 4,062-strong trust and safety team was cut down to 2,849, while its full-time content moderation team was cut in half, from 107 down to 51, though most of X's moderators are contractors. The company had 2,613 contracted moderators in October 2022, which fell to 2,305 in May 2023.

The cuts have had an impact, according to eSafety, which said users reported response times to hateful tweets had slowed by 20% since Musk's takeover. Company attempts to deal with hateful direct messages had slowed by 70%, the government agency said.

X told the commissioner that after [disbanding its Trust and Safety Council](#), which provided external advice to the company, it had not sought to establish a replacement body. X also confirmed to the agency it didn't have any full-time staff singularly dedicated to hateful conduct issues globally, either before or after Musk took over.

Some Community Notes contributors (who are also NAFO activists) have claimed that the system is riddled with coordinated manipulation, infighting and lack of oversight from the platform. Also, many contributors engage in conspiracy-fueled discussions.

8/15

"I understand why they do it, but it doesn't do anything like what they say it does," one Community Notes contributor tells WIRED. "It is prone to manipulation, and it is far too slow and cumbersome. It serves no purpose as far as I can see. I think it's probably making the disinformation worse, to be honest."

This contributor, who was granted anonymity for this story because of concerns about retaliation from Musk or his supporters, should know about platform manipulation, because he is one of those engaging in it.

He is one member of a group of about 25 Community Notes contributors who actively coordinate on a daily basis in order to upvote or downvote particular notes. He is a member of the North Atlantic Fella Organization (NAFO), a decentralized online volunteer movement that emerged following Russia's invasion of Ukraine and uses humor to counter Russian disinformation and propaganda while raising funds in support of Ukraine.

"We have a group so we can coordinate who writes what Community Notes, and if you write one, you can coordinate so you get one voted as helpful, and there's no validation on any of that," the contributor says. The contributor claims that Russian embassies have engaged in similar tactics and alleges that they "will downvote anything that's anti-Russian." A WIRED review of unpublished notes accompanying contested Russian embassy posts written by NAFO members shows they have all been repeatedly downvoted.

The Notes system also has a huge problem with its scalability. During events like the 7 Oct 2023 Hamas terrorist attack, the amount of disinformation grows so large, that it's simply impossible for the small community to keep up and check factuality of said content.

9/15


Mashable Super Bowl 2024 AI @ WORK Tech Science Life Social Good Entertainment Deals Shopping

Home > Tech

Most users on X never see Community Notes correcting misinformation

A Mashable investigation found that Elon Musk's answer to X's misinformation problem is failing.

By [Matt Binder](#) on November 30, 2023 [f](#) [X](#) [v](#)



Elon Musk's X depends on Community Notes to fact-check misinformation on the platform. A Mashable investigation found that most users who view the misinformation don't see those notes. Credit: Jakub Porzycki/NurPhoto via Getty Images

Analysis by NewsGuard showed that the most popular disinformation posts related to the Israel-Hamas war (not so surprisingly originating from serial liars like @jacksonhinkle, @drloupis and @ShaykhSulaiman) failed to receive Community Notes 68% of the time.

10/15



HOME > DIGITAL > NEWS

Oct 20, 2023 7:22am PT

X/Twitter Verified Blue Check-Mark Users Are 'Superspreaders' of Disinformation About Israel-Hamas War, Study Says

By Todd Spangler ▾



X Corp.

These big accounts also have the ability to fight against the Notes they've received by mobilizing people who support their views. In the most tragicomic instance, @elonmusk claimed, without any evidence, that a Community Note on his post was "gamed by state actors".



Jackson Hinkle 🇺🇸

@jacksonhinkle

Subscribe



URGENT: RATE THIS COMMUNITY NOTE “UNHELPFUL.”



@haaretzcom’s source was within the Israeli Police. How is that “not credible?”
Community notes have been HIJACKED BY ZIONISTS!



Elon Musk 🇺🇸

@elonmusk

Subscribe

Wikipedia is inherently hierarchical and therefore subject to the biases of higher ranking editors, independent of their merits.

@CommunityNotes requires people with historically different points of view, based on how they have rated and written Notes, to agree in order for Notes to be shown to the public. All code and data is open source, so you can recreate the outcome yourself.

Crucially, even I, as the controlling shareholder of the company, cannot change the outcome of a Note.

This is an extremely fundamental difference.



Elon Musk @elonmusk



An American YouTuber has been imprisoned for 5 years in Ukraine, and allegedly tortured, for making videos.

This is not ok. [twitter.com/tuckercarlson/...](https://twitter.com/tuckercarlson/)



Tucker Carlson @TuckerCarlson · Dec 10 ·

Ep. 47 Gonzalo Lira is an American citizen who's been tortured in a Ukrainian prison since July, for the crime of criticizing Zelensky. Biden officials approve of this, because they'd like to apply the same standard here. The media agree. Here's a statement from Gonzalo Lira's...

[Show more](#)



Readers added context →

Gonzalo Lira was arrested in May for justification/denial of the Russian invasion of Ukraine, a crime under Article 436-2 of the Criminal Code, not for "criticizing Zelensky". He was released on bail but tried to leave the country. There is no evidence of torture.

- justice.gov/sites/default/...
- t.me/SBUkr/8224
- news.yahoo.com/pro-russian-bl...
- opendatabot.ua/court/11177088...
- news.yahoo.com/sbu-responds-m...
- thedailybeast.com/gonzalo-lira-r...

Readers added context they thought people might want to know →

Gonzalo Lira was arrested on May 1st and is currently being held in the Kharkiv pretrial detention center. His court hearing is scheduled for Dec. 12 and Dec. 21. He has not "been imprisoned for 5 years in Ukraine" nor is there any evidence of him having been tortured.

- news.yahoo.com/pro-russian-bl...
- opendatabot.ua/court/11177088...
- news.yahoo.com/sbu-responds-m...
- thedailybeast.com/gonzalo-lira-r...

 **Elon Musk**   @elonmusk · 15h ...

What is the status of this American journalist [@JoeBiden](#)?

 **Gonzalo Lira** @GonzaloLira... · 7/31/23

Right now, I'm about to try to get out of Ukraine, and seek political asylum in Hungary.

Either I'll cross the border and make it t...

 **Readers added context**

Lira is actually a Youtuber. He has violated Article 436-2 of Ukraine's Criminal Code.

He denied the Bucha massacre and othe...

[Show this note](#)

 7,914  18.2K  78.3K  22.6M 

 **Elon Musk**   @elonmusk · 8h ...













Interesting. This Note is being gamed by state actors. Will be helpful in figuring who they are.














Thanks for jumping in the honey pot, guys
lmao!

 3,097  3,626  19.2K  2.3M 

Other than being humiliated and ridiculed, getting Community Noted doesn't really have any major downsides. Noted posts don't provide you income and advertisers can decide if they want to show ads on accounts like @dom_lucre's, but most of these...

Community Notes Leaderboard

	Rank	Notes
 CHOQUEI @choquei	#1	357
 illuminatibot @illuminatibot	#2	292
 Jackson Hinkle @jacksonhinkle	#3	276
 Daily Loud @DailyLoud	#4	227
 Ian Miles Cheong @stillgray	#5	191
 Dom Lucre Breaker of Narratives @dom_lucre	#6	171
 Sulaiman Ahmed @ShaykhSulaiman	#7	143
 Stew Peters @realstewpeters	#8	140
 Dr. Anastasia Maria Loupis @DrLoupis	#9	136
 Concerned Citizen @BGatesisaPsycho	#10	128
 Mario Nawfal @MarioNawfal	#11	126
 @Sprinter0000 @Sprinter0000	#12	123
 Laughing Man @jhdrei	#13	120

	Dexerto @Dexerto	#14	117
	Laura Loomer @LauraLoomer	#15	116
	MFA Russia @mfa_russia	#16	111
	Visegrád 24 @visegrad24	#17	107
	DramaAlert @DramaAlert	#18	100
	Wall Street Silver @WallStreetSilv	#19	92
	Matt Wallace @MattWallace888	#20	89
	World of Statistics @stats_feed	#21	84
	ツイッター速報 BreakingNews @tweetsoku1	#22	79
	DR. Kek @Theekesociety	#23	76
	POPTime @siteptbr	#24	70
	Russia in South Africa @EmbassyofRussia	#25	68
	Elon Musk @elonmusk	#26	65

...so-called superspreader accounts make most of their income through other means, namely through the X's subscription system. Also, many of them, including @stillgray and (allegedly) @jacksonhinklle are employed by state actors like Russia and the CCP.

TWITTER · X / TECH / POLITICS

Blue checkmarks on X are 'superspreaders of misinformation' about Israel-Hamas war



X's community notes feature isn't doing much to prevent the spread. Illustration: The Verge

/ A new study finds that most of the 'false or unsubstantiated narratives' go viral on X before spreading to other platforms, with verified users responsible for 74 percent of the most engaged content.

By [Jess Weatherbed](#), a news writer focused on creative industries, computing, and internet culture. Jess started her career at TechRadar, covering news and hardware reviews.

Oct 20, 2023, 3:01 PM GMT-3



RT

QUESTION MORE

LIVE

[Russia & Former Soviet Union](#)

[World News](#)

[Business](#)

[Sport](#)

[RT Features](#)

[/ Home](#) / [/ Tags](#) / [/ Op-ed](#) / [/ All authors](#)

Ian Miles Cheong



***Ian Miles Cheong** is a political and cultural commentator. His work has been featured on [The Rebel](#), [Penthouse](#), [Human Events](#), and [The Post Millennial](#). Follow Ian on Twitter [@stillgray](#) and on Telegram [@CultureWarRoom](#)*

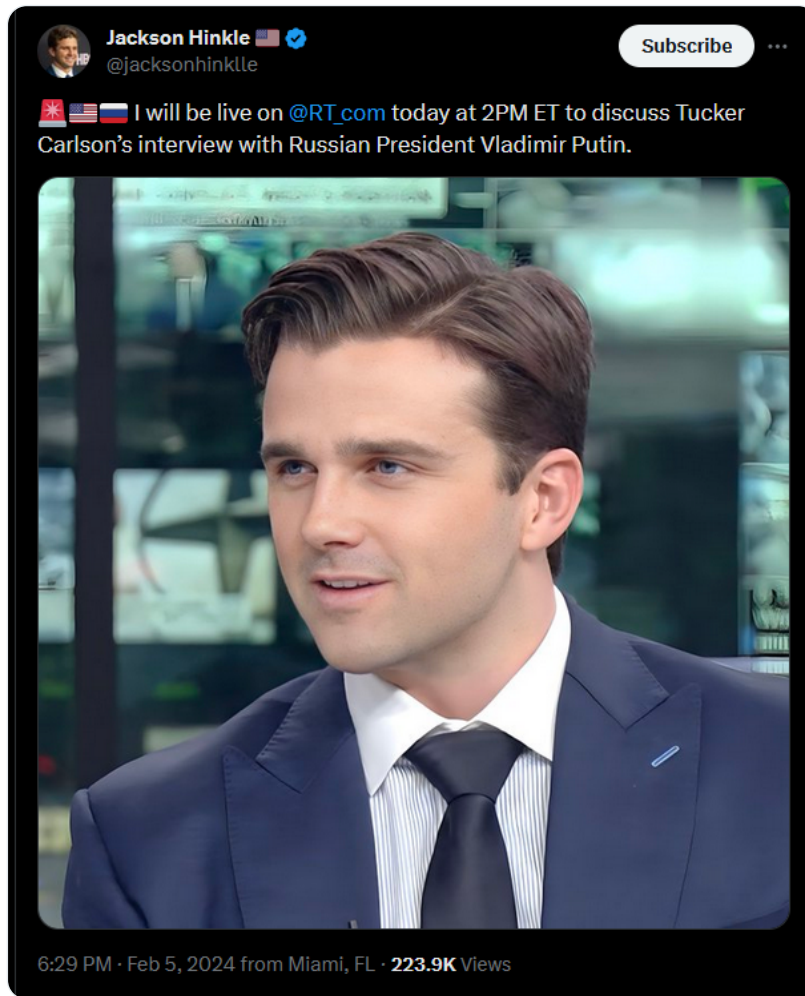


Jackson Hinkle 🇺🇸 ✓
@jacksonhinklle

Subscribe

🇺🇸🇨🇳🇺🇸 Speaking on China's #1 news today about war, Communism & multipolarity.





With accounts that post tens or hundreds posts a day, the Notes are also inefficient - while the Community is trying to put a note on a post that's clearly disinformation, there are already 10 or 20 new ones to replace it in the algorithm.

14/15

How Musk's X Is Failing To Stem the Surge of Misinformation About Israel and Gaza

By [Davey Alba](#), [Denise Lu](#), [Leon Yin](#) and [Eric Fan](#)
for **Bloomberg Technology**
November 21, 2023

[f](#) [t](#) [in](#) [e](#) [envelope](#) Gift this article

To conclude, Community Notes are a non-functional and slow mechanism that's desperately trying to replace the Trust and Safety team. They work on a "mob rule" basis and big enough

accounts (including the owner of the platform) can play around the system.

15/15

Search Q

February 8, 2024

Poynter. NEW

NEWS TRAINING ETHICS & LEADERSHIP FACT-CHECKING MEDIA LITERACY EVENTS CONTEST

Fact-Checking IFCN Media Literacy

Why Twitter's Community Notes feature mostly fails to combat misinformation

An attempt to solve 'the problem of biased fact-checkers' has left most Notes unseen by public, fact-checking conference is told



MediaWise director Alex Mahadevan talks about the flaws in Twitter's Community Notes system at GlobalFact 10 in Seoul, South Korea

By: [Madison Czopek](#)

June 30, 2023

f t p e

Start your day informed and inspired.

SUBSCRIBE

Get the Poynter newsletter that's right for you.

Media Jobs

Newsletter Editor - Seattle, WA (98122)
Cascade PBS
Classical Music Radio Host/Producer - Pullman, WA (99164)

All soups:

Find us also on other socials:



Vatnik Soup

#vatniksoup is a Twitter thread series (and a hashtag!) where I'll introduce pro-Russian actors and propagandists from around the world, be they so-called "independent journalists", politicians, mili...

<https://www.vatniksoup.com>

[instagram.com/vatniksoup/](https://www.instagram.com/vatniksoup/)



The Soup Central

The Soup Central gives you in-depth knowledge on disinformation, bad actors and information operations.

<https://www.youtube.com/@TheSoupCentral/videos>



...