Twitter Thread by Not Chris Nandor

Not Chris Nandor
@pudge_nandor

From @jack:

Starting now, we are changing the way we do things at Twitter. Even though we have been following policies we created, that we thought were good and necessary, it just hasn't had the desired effect of promoting good discourse. Effective immediately, we are changing.

The basic principle that now guides us is that instead of trying to manipulate discourse, we are opening it up, and making ourselves more transparent and accountable, and resisting attempts at censorship of us or our users.

Further, we have been fundamentally operating under the false premise that we, our our news and fact-checking partners, are smarter or better at understanding information than you, our users, are. This is a vain fiction, and we will drop this pretense.

We have long known that we are not just another social media site, but the place where citizens, governments, and organizations communicate with each other, for essential purposes. We have a great responsibility to the public to be open. So, we change to improve.

First: we will no longer proactively delete any content, except for: private information harming a person other than who posted it (e.g. doxxing, revenge porn, etc.), and any content we are legally required to remove.

That's it. There are, for now, no other categories. I hope there will not be. We will continue to have a close working relationship with the DOJ and other law enforcement agencies so we can respond very quickly to their requests, but they will initiate removal, not us.

And when we remove content, it will all be documented clearly and available for auditing by independent third parties (as many as wish to do so), who will sign NDAs agreeing to not disclose the removed content, but will be free to report on all other aspects of it.

Second: we will continue to suspend accounts for bad behavior, but we will be entirely transparent about it. When you go to an account that is suspended, you will see that it is suspended, and the content that got the account suspended, and the

policy it violated.

Obviously, for removed content, we will not show the content, but will show a placeholder and the policy violated, and the third-party independent auditors will be able to see the content, and report for themselves whether it was justified.

Third: we will have a serious appeals process. No one appeals now because there is no point: we never overturn decisions except by public pressure. But appeals require resources. So, starting one week from today, we will offer subscriptions on Twitter.

For $1/mo., you will get no ads, and expedited appeals. These expedited appeals will happen within one hour, any time of day or night. If your appeal takes longer than an hour, it is automatically upheld.

If your appeal is upheld, you will get a free subscription for one year, so we are incentivized to get to the appeal quickly, and to get the action right the first time.

We have internal tools that rate the ideology of our users in multiple ways. We will use these tools to score our employees, so that if they are taking action against users or posts in an ideologically lopsided manner, we will detect it and take action.

Any employee found to be repeatedly violating these policies will be reassigned or released. The independent third-party auditors will have full access — with personal information removed, and employees identified by random ID — to this history.

Fourth: we will no longer engage in any fact-checking. Our fact-checkers — and the fact-checkers at WaPo and other news organizations — are no better at fact-checking than most of you are. It is a waste of time and resources and we often do it in a biased way.

Fifth: we will no longer take sides on any issues. Our role is to promote discourse, not to steer it. We will continue to summarize news, but we will not adopt a viewpoint on the news items.

Sixth: we will offer new tools to promote a good user experience. Just because we are going to allow someone to post anti-Semitic views, doesn't mean you should have to see it.

We will use algorithms to hide content from your view, unless you specifically request to see it, similar to how mutes work today. A key component of this algorithm is the new Dislike feature. Posts that are disliked significantly are hidden.

Your social network graph affects this algorithm: accounts you follow will not be hidden; accounts they follow will be less likely to be hidden. And so on.

We will also fill the space between blocks and mutes: "Ignore" will be like many think "mute" should work: you will never, ever see that account's content (unless you request it), but they can still see you. When they view your account, they will see that they are being Ignored.

We will also allow you to block not just keywords, but concepts and ideologies. Block discussions about sex, or religion. Or block accounts we have marked as anti-Semitic. The categories and data used to populate them will be fully auditable by the independent third parties.

This last one is very scary to us, but we have the tools to do it, and it will be transparent and accountable. We believe it is a good balance between a good user experience, and being open.

Some of you will be disheartened by this. You want us to silence speech that you dislike … usually for good reasons. But we learned — and we should have known all along — that there is no reasonable path to do so, because there are no people capable of reasonably enforcing it.

So we will stop trying to silence anyone, and instead focus on a great user experience for all of our users, who are mostly just trying to connect with other people, and we will support that, regardless of ideology.

— @jack