Hey folks —

Wanted to flag the below tweet and am wondering if we can get moving on the process for having it removed ASAP:

https://twitter.com/RobertKennedyJr/status/135274813965645569

And then if we can keep an eye out for tweets that fall in this same ~genre that would be great.

Thanks!

Clarke
Thanks. We recently escalated this.

On Fri, Jan 22, 2021 at 8:05 PM Humphrey, Clarke EOP/WHO [redacted]@who.eop.gov> wrote:

Hey folks —

Wanted to flag the below tweet and am wondering if we can get moving on the process for having it removed ASAP:

>https://twitter.com/RobertKennedyJr/status/1352748139656455699<

And then if we can keep an eye out for tweets that fall in this same ~genre that would be great.

Thanks!

Clarke
Hi Rob,

Glad that we could help resolve the issue last night. To help streamline the process, and ensure that you have expedited help, we would strongly recommend the following:

**1. Consult with the White House IT Department to unblock emails from Twitter’s Support Ticketing System.** The issues you’re experiencing are due to the White House’s system prohibiting emails. The two prior administrations also experienced this issue and it is fixable within your internal systems. This is particularly critical to resolve at large because if there is an issue with your account, we would notify you through email.

**2. Designate a list of authorized White House staff for Twitter’s Partner Support Portal.** We sent over instructions about this on January 28th and also discussed this with Christian during our call on February 4th. This is the same system we had in place for the previous two administrations for their support issues, as well as the transition and campaign teams.

Once you assign and we enroll these authorized reporters, whenever they submit a ticket through the Help Center it will be prioritized automatically, without having to contact our team, and you won’t need to add your personal information. To enroll your designated reporters to the Partner Support Portal, we simply need the list of @usernames (up to 10) that are registered with a White House email address.

**3. Streamlined coordination with ODS.** We are committed to making sure your team is properly trained and equipped with all of the tools and best practices for both content development and triaging issues. To deliver the best service, we would prefer to have a streamlined process strictly with your team as the internal liaison. That is the most efficient and effective way to ensure we are prioritizing requests. In a given day last week for example, we had more than four different people within the White House reaching out for issues. The more we can empower your team to be the in-house experts, the better service and partnership we can provide.

I would welcome a conversation about the aforementioned if you have specific questions.

Thanks,

Twitter, Inc. | Public Policy
@TwitterGov & @Policy

On Sat, Feb 6, 2021 at 11:09 PM Flaherty, Robert EOP/WHO wrote:

Thanks

Sent from my iPhone

On Feb 6, 2021, at 10:32 PM, [redacted] wrote:
Update for you - account is now suspended.

On Sat, Feb 6, 2021 at 9:47 PM Flaherty, Robert EOP/WHO @who.eop.gov wrote:
Great. Cannot stress the degree to which this needs to be resolved immediately.

Sent from my iPhone

On Feb 6, 2021, at 9:47 PM, @twitter.com wrote:
Thank you for sending over. We’ll escalate for further review from here.

On Sat, Feb 6, 2021 at 9:45 PM Flaherty, Robert EOP/WHO @who.eop.gov wrote:
I have tried using your form three times and it won’t work — it is also ridiculous that I need to upload my id to a form prove that I am an authorized representative of Finnegam Biden.

Please remove the is account immediately:

>>>https://twitter.com/bidenfinnegan<<<;

I have CC’d Anthony Bernal, the First Lady’s senior advisor, in case you have any further questions.

Sent from my iPhone

--

Twitter | Public Policy
--

Twitter | Public Policy
Hi Rob,

Quickly following up to see when you would like to have a meeting arranged to speak to our misinformation team reps about the latest updates. They also have a more detailed misinformation analysis prepared based on the discussions/questions from the previous meetings during the transition time period.

Best,

Get Outlook for iOS

Good evening Rob,

We have provided responses to your initial questions with input from the various teams below. We are happy to discuss these and additional questions as per your recent note. Do let us know a few windows that work for you.

Can you share more about your framework here? May, of course, is very different than “will.” Is there a strike policy, ala Youtube? Does the severity of the claims matter?

We don’t disclose the details of our thresholds publicly due to concerns about users gaming the system to avoid enforcement, however we do notify Groups, Pages, and Advertisers when we’ve removed content that violates our Community Standards. We start placing restrictions on accounts, Pages, and Groups for multiple violations, including restrictions on their ability to share content for increasing periods of time and limitations on their ability to reach their audience. If violations continue, we will suspend the entire Page, Group, or account. Additionally, when we review Pages and Groups we look at how they describe themselves and may restrict or remove them if the title or description violate our policies.

And as far as your removal of claims, do you have data on the actual number of claims-related posts you’ve removed? Do you have a sense of how many are being flagged versus how many are being removed? Are there actions (downranking, etc) that sit before removal?

It is a bit too early to be sure - We will begin enforcing this policy immediately, with a particular focus on Pages, Groups and accounts that violate these rules, and we’ll continue to expand our enforcement over the coming weeks. There is a
range of content that can violate these policies, and it will take some time to train the reviewers and systems on enforcement.

How are you handling things that are dubious, but not provably false?

In consultation with leading health organizations, we continuously expand the list of false claims that we remove about COVID-19 and vaccines during the pandemic. We remove claims public health authorities tell us have been debunked or are unsupported by evidence.

Content which does not qualify for removal may be eligible to be fact-checked by our network of over 80 fact-checking organizations. When one of our independent fact-checking partners debunk a post, we reduce its distribution and add strong warning labels with more context, so fewer people see the post. We do not remove the content, but are focusing on improvement efforts that will help us to better address content that contributes to unfounded hesitancy towards the COVID-19 vaccine.

For example, we’re working to proactively prevent posts discouraging vaccines from going viral on our platforms; address content that experts believe dissuades people from getting the vaccine, but does not violate our misinformation policies, through the use of information labels; and prevent recommendations for Groups, Pages, and Instagram accounts that repeatedly push content discouraging vaccines.

-On Behalf of the Facebook team

FACEBOOK

U.S. Public Policy
Facebook

From: "Flaherty, Robert EOP/WHO" flaherty.re@who.eop.gov
Date: Tuesday, February 9, 2021 at 4:59 PM
To: flaherty.re@who.eop.gov, "Rowe, Courtney M. EOP/WHO" rowe.cs@who.eop.gov,
"Humphrey, Clarke EOP/WHO" humphrey.ch@who.eop.gov
Cc: flaherty.re@who.eop.gov, rowe.cs@who.eop.gov, humphrey.ch@who.eop.gov
Subject: RE: COVID-19 Outreach to communities worldwide

All, especially given the Journal’s reporting on your internal work on political violence spurred by Facebook groups, I am also curious about the new rules as part of the “overhaul.” I am seeing that you will no longer promote civic and health related groups, but I am wondering if the reforms here extend further? Are there other growth vectors you are controlling for?

Happy to put time on the calendar to discuss further.

From: Flaherty, Robert EOP/WHO
Sent: Monday, February 8, 2021 1:37 PM
To: flaherty.re@who.eop.gov, Rowe, Courtney M. EOP/WHO rowe.cs@who.eop.gov, Humphrey, Clarke EOP/WHO humphrey.ch@who.eop.gov
that repeatedly share these debunked claims may be removed altogether.

Can you share more about your framework here? May, of course, is very different than “will.” Is there a strike policy, ala Youtube? Does the severity of the claims matter?

And as far as your removal of claims, do you have data on the actual number of claims-related posts you’ve removed? Do you have a sense of how many are being flagged versus how many are being removed? Are there actions (downranking, etc) that sit before removal? How are you handling things that are dubious, but not provably false?

Thanks

Good afternoon Courtney, Rob, and Clarke,

We wanted to make sure you saw our announcements today about running the largest worldwide campaign to promote authoritative COVID-19 vaccine information and expanding our efforts to remove false claims on Facebook and Instagram about COVID-19, COVID-19 vaccines and vaccines in general during the pandemic. More details are in our Newsroom: authoritative COVID-19 vaccine information and COVID-19 and vaccine misinformation.

Helping People Find Where and When They Can Get Vaccinated

• Starting this week, we’ll feature links in the COVID-19 Information Center to local ministry of health websites to help people understand whether they’re eligible to get vaccinated and how to do so.
• And in the coming weeks, as more information becomes available, we’ll continue to improve this feature, making it easier for people to see where and when they can get vaccinated in just a few taps.

Sharing Credible Information About COVID-19 Vaccines

• We’re working with health organizations and community leaders to run campaigns on our platform promoting accurate information about COVID-19 vaccines and encouraging people to get vaccinated.
• We’re giving over $120 million in ad credits to help health ministries, NGOs and UN agencies reach billions of people around the world with COVID-19 vaccine and preventive health information.
• In the US, we’re partnering with the Johns Hopkins Bloomberg School of Public Health to reach Native American communities, Black communities and Latinx communities, among others, with science and evidence-based content that addresses the questions and concerns these communities have.
• We’re also working with AARP to reach Americans over 50 with educational content about COVID-19 vaccines, including Spanish-language content designed to reach Latinx and Hispanic communities.

Combating Vaccine Misinformation
• We are expanding our efforts to remove false claims on Facebook and Instagram about COVID-19, COVID-19 vaccines, and vaccines in general during the pandemic. Since December, we’ve removed false claims about COVID-19 vaccines that have been debunked by public health experts.

• Today, following consultations with leading health organizations, including the World Health Organization (WHO), we are expanding the list of false claims we will remove to include additional debunked claims about the coronavirus and vaccines. We already prohibit these claims in ads.

• Groups, Pages and accounts on Facebook and Instagram that repeatedly share these debunked claims may be removed altogether. We are also requiring some admins for groups with admins or members who have violated our COVID-19 policies to temporarily approve all posts within their group.

• When people search for vaccine or COVID-19 related content on Facebook, we promote relevant, authoritative results and provide third-party resources to connect people to expert information about vaccines. On Instagram, in addition to surfacing authoritative results in Search, in the coming weeks we are making it harder to find accounts in search that discourage people from getting vaccinated.

• As we noted last month in response to guidance from the Oversight Board, we are committed to providing more transparency around these policies. You can read the detailed updates in Facebook’s Community Standards and in our Help Center.

Providing Data to Inform Effective Vaccine Delivery

• Last year, we began collaborating with Carnegie Mellon University Delphi Research Group and the University of Maryland on COVID-19 surveys about symptoms people are experiencing, mask wearing behaviors and access to care. With over 50 million responses to date, the survey program is one of the largest ever conducted and has helped health researchers better monitor and forecast the spread of COVID-19.

• To help guide the effective delivery of COVID-19 vaccines, the survey data will provide a better understanding of trends in vaccine intent across sociodemographics, race, geography and more. The scale of the survey will also allow for faster updates on changes in trends, such as whether vaccine intent is going up or down in California in a given week and better insights on how vaccine intent varies at a local level. We’ll share these new insights including vaccine attitudes at a county level in the US as well as globally.

These new policies and programs will help us continue to take aggressive action against misinformation about COVID-19 and vaccines and help people find where and when they can get vaccinated. You can read more about how we’re supporting COVID-19 relief efforts and keeping people informed at our COVID-19 action page.

-On Behalf of the Facebook team

FACEBOOK

U.S. Public Policy
Facebook
thanks. Good insights here.

I'm more interested in the data that was outlined in the Washington Post
(https://www.washingtonpost.com/technology/2021/03/14/facebook-vaccine-hesitancy-qanon)

And what interventions you are testing/their effectiveness.

-Rob

Sent from my iPhone

On Mar 12, 2021, at 4:59 PM, <fb.com> wrote:

Hi All,

Following up on our commitment to share our survey data on vaccine uptake. We're happy to share these findings regularly moving forward to help inform your teams and strategies. Attached are our findings from January 10 -- February 27, 2021. On Monday the report will be available online, and I'll be sure to send a link when it's published.

Note that highlights of the findings are up top, a robust executive summary follows, and then a deep dive into the methodology, greater detail on state trends, occupations, barriers to acceptance etc. Hopefully, this format works for the various teams and audiences within the White House / HHS that may find this data valuable.

We're also open to feedback on the formatting.

Please let us know if you have specific questions about the findings or the survey itself, we're happy to track down answers or book time.

Best,

---

facebook, inc. politics & government
<fb.com>
Thanks Andy, and apologies for the delay in getting back. We are absolutely invested in getting you the specific information needed to successfully manage the vaccine rollout. We want to share information with you that we trust is statistically significant and derived from sound analysis, so that it can actually be helpful. The information cited in the WaPo article over the weekend was leaked and was not vetted internally to understand how accurate it is or the ramifications that could result from it. But I understand your point regarding how we communicate, and that we need to share information with you in a way that prioritizes what we are seeing in as close to real time as possible. I’d like to set up a conversation with our research leads to walk your team through ongoing research we are currently conducting and our approach; and then we can prioritize sharing results as quickly as possible.

Moreover, the data we sent on Friday and will continue to send throughout the year represents the information we are using internally to shape our own thinking on this content—we believe this data addresses many of the questions that have been posed (because it has been so helpful to guide our own internal efforts). We’d appreciate the opportunity to go through it in detail with whomever is interested on your team.

I know you’re extremely busy. If it’s ever helpful to connect by phone instead of over email I am at [redacted].

---

I appreciate being copied on the note. It would nice to establish trust. I do feel like relative to others, interactions with Facebook are not straightforward and the problems are worse — like you are trying to meet a minimum hurdle instead of trying to solve the problem and we have to ask you precise questions and even then we get highly scrubbed party line answers. We have urgency and don’t sense it from you all. 100% of the questions I asked have never been answered and weeks have gone by.

Internally we have been considering our options on what to do about it.

Regards,

Andy

Sent from my iPhone

On Mar 15, 2021, at 6:42 PM, [redacted]@fb.com> wrote:
Thanks, Rob. Called and left you a message earlier. I understand why you'd read the WaPo piece and come away feeling like we are not leveling with you. The piece inflated unconfirmed and leaked work that's being done by a small team. It's exploratory work and is not close to being a finalized work product - in fact the team that briefed you (including me) wasn't aware of the work at the time we briefed you. This was not a “massive study” as depicted by the Post - this was a small team experimenting with applying a relatively new system to COVID19 content. At any given time, there are many research projects similar to this being conducted by data scientists across the platform -- as we've discussed, we're working hard to understand and address this type of content. Our definition of vaccine hesitancy is evolving - it is not a mature concept. This is early work and we have not gone through the kind of quality assurance we'd usually do before sharing the learnings externally. The data that leaked and was reported on should not be interpreted to be anything more than one of many efforts underway to better inform how we tackle this problem. As we develop them further, we will definitely keep you updated.

We obviously have work to do to gain your trust. You mention that you are not trying to play “gotcha” with us — I appreciate the approach you are taking to continued discussions. We are also working to get you useful information that's on the level. That's my job and I take it seriously — I'll continue to do it to the best of my ability, and I'll expect you to hold me accountable.

If interested, I can schedule time to give you more context on how this work is done and why we wouldn't include it in a briefing.

From: Flaherty, Rob EOP/WHO
Date: Monday, March 15, 2021 at 1:10 PM
To: [email protected]
Cc: Slavitt, Andrew M. EOP/WHO
Subject: RE: You are hiding the ball

I don't think this is a misunderstanding, I've been asking you guys pretty directly, over a series of conversations, for a clear accounting of the biggest issues you are seeing on your platform when it comes to vaccine hesitancy, and the degree to which borderline content - as you define it - is playing a role. I've also been asking for what actions you have been taking to mitigate it as part of your “lockdown” - which in our first conversation, was said to be in response to concerns over borderline content, in our 1:1 conv you said was not out of any kind of concern over borderline content, and in our third conversation never even came up.

You said you would commit to us that you'd level with us. I am seeing in the press that you have data on the impact of borderline content, and its overlap with various communities. I have asked for this point blank, and got, instead, an overview of how the algorithm works, with a pivot to a conversation about profile frames, and a 45-minute meeting that seemed to provide you with more insights than it provided us.

I am not trying to play “gotcha” with you. We are gravely concerned that your service is one of the top drivers of vaccine hesitancy - period. I will also be the first to acknowledge that borderline content offers no easy solutions. But we want to know that you're trying, we want to know how we can help, and we want to know that you're not playing a shell game with us when we ask you what is going on.

This would all be a lot easier if you would just be straight with us.

From: [email protected]
Sent: Monday, March 15, 2021 10:22 AM
To: Flaherty, Rob EOP/WHO
Cc: Slavitt, Andrew M. EOP/WHO
Subject: [EXTERNAL] Re: You are hiding the ball
Thanks Rob—I think there is a misunderstanding on what this story is covering with respect to research that’s happening—I will call to clear up. Certainly not hiding the ball.

Also flagging our announcement that went live this morning—this is the announcement I mentioned on Friday’s call.


From: Flaherty, Rob EOP/WHO  
Date: Sunday, March 14, 2021 at 11:13 PM 
To:  
Cc: Slavitt, Andrew M. EOP/WHO  
Subject: You are hiding the ball

>>>https://www.washingtonpost.com/technology/2021/03/14/facebook-vaccine-hesitancy-qanons<<<;

Sent from my iPhone
Look forward to talking today at 4:00. I will plan on giving an overview of her role and the work across the teams at the top and of course will respond to questions, as that’s the objective of having her in touch with you regularly over the coming weeks. One additional participant on our end will be — just to make sure we’re tracking all follow ups.

Great. I can do 4!

Rob—we’re good to schedule around your avail Wednesday afternoon if that works.

---

---

---

Thanks Rob—appreciate the context below. For the meeting with — possible that we could aim for Wednesday? I’ll rally our folks if you have a window in the afternoon that will work.
Awesome. Similarly to how we’re looking out for your game plan on tackling vaccine hesitancy spread on your platform, we’ll look out for how you plan to help close the gap on equitable access.

Had a chance to connect with Andy earlier to download on his call with — seems like there’s alignment here.

Excited to meet — Could talk tomorrow in the 4-5 hour ET tomorrow.

As for sharing data, that’s great. Again, as I’ve said, what we are looking for is the universe and scale of the problem. You noted that there is a level below sensational stories that get down-ranked, which took the form of general skepticism. I think it is helpful to know where you think the biggest issues is. I think we are all aligned that the problem does not sit in “microchips”-land, and that it seems plausible that the things that drive the most actual hesitancy sit in “sensational” and “skeptical.” If you’re down-ranking sensational stuff—great—but I want to know how effective you’ve seen that be from a market research perspective. And then, what interventions are being taken on “skepticism?” I could see a range of actions, including hitting them good information, boosting information from sources they’ve indicated they trust, promoting content from their friends who have been vaccinated…….what are you trying here, and again, how effective have you seen it be. And critically, what amount of content is falling into all of these buckets? Is there wider scale of skepticism than sensationalism? I assume given the Carnegie data and the studies I’ve seen in the press that you have this. While I think you and I both know that access to the study’s toplines and a crowdtable account aren’t going to get us the info we’re looking for, it shows to me that you at least understand the ask.

As I’ve said: this is not to play gotcha. It is to get a sense of what you are doing to manage this. This is a really tricky problem. You and I might disagree on the plan, but I want to get a sense of the problem and a sense of what you solutions are.

On whatsapp, which I may seem like I’m playing gotcha, but I guess I’m confused about how you’re measuring reduction of harm. If you can’t see the message, I’m genuinely curious—how do you know what kinds of messages you’ve cut down on? Assuming you’ve got a good mousetrap here, that’s the kind of info we’re looking for above: what interventions you’ve taken, and what you’ve found to work and not work? And how effective are you seeing the good information on WhatsApp be? Are you doing cross-platform campaign work to try to reduce people’s exposure on Whatsapp? As we worry about equity and access, Whatsapp is obviously a central part of that given its reach in immigrant communities and communities of color.

You’ve given us a commitment to honest, transparent conversations about this. We’re looking for that, and hoping we can be partners here, even if it hasn’t worked so far. I know Andy is willing to get on the phone with a couple of times per week if its necessary to get all of this.

Looking forward.
Thanks Andy. Also—wanted to flag a discussion we are scheduled to have with regarding some work around equitable vaccine adoption—just a touch-base conversation to talk through ideas we have for closing the adoption gap in communities disproportionately impacted by Covid and to discuss how we can be supportive overall in the US re: an equity strategy. We were connected with who scheduled the conversation—just didn’t want any surprises.

From: Slavitt, Andrew M. EOP/WHO @who.eop.gov
Date: Monday, March 22, 2021 at 9:37 AM
To: @fb.com>
Cc: Flaherty, Rob EOP/WHO @who.eop.gov>
Subject: RE: Follow up - Friday call with

Thanks and I will connect and follow up.

From: @fb.com>
Sent: Sunday, March 21, 2021 11:25 PM
To: Slavitt, Andrew M. EOP/WHO @who.eop.gov>
Cc: Flaherty, Rob EOP/WHO @who.eop.gov>
Subject: [EXTERNAL] Follow up - Friday call with

Andy,

Thanks for taking the time to connect on Friday. Per our discussion, I wanted to follow up with next steps:

1. **Consistent Product Team POC:** As discussed, we will make who has been coordinating the product work that matters most to your teams, available on a regular basis. If it makes sense, we can schedule some time for to connect with you and/or Rob (and whomever else makes sense) early this week.
2. **Sharing Additional Data:** Mentioned the new internal analytics that we are developing to help us understand and monitor the most viral COVID vaccine-related content. This is a top priority for us, and we will keep you updated on our progress and when we expect to be able to share the data with you.
3. **Levers for Tackling Vaccine Hesitancy Content:** You also asked us about our levers for reducing virality of vaccine hesitancy content. In addition to policies previously discussed, these include the additional changes that were approved late last week and that we’ll be implementing over the coming weeks. As you know, in addition to removing vaccine misinformation, we have been focused on reducing the virality of content discouraging vaccines that does not contain actionable misinformation. This is often-true content, which we allow at the post level because experts have advised us that it is important for people to be able to discuss both their personal experiences and concerns about the vaccine, but it can be framed as sensational, alarmist, or shocking. We’ll remove these Groups, Pages, and Accounts when they are disproportionately promoting this sensationalized content. More on this front as we proceed to implement.
4. **WhatsApp:** Finally—mentioned the policies that apply to WhatsApp. WhatsApp’s approach to misinformation focuses on limiting the virality of messages, preventing coordinated abuse, and empowering users to seek out reliable sources of information both in and out of the product. Our product includes features to limit the spread of viral content, such as forward limits and labels, privacy settings to help users decide who can add them to groups, and simple ways for users to block accounts and make reports to WhatsApp if they encounter problematic messages. Additional limitations we placed in April 2020 on forwarding of messages that have been forwarded many times reduced these kinds of messages by over 70%.

Along with these commitments, we’ll continue to provide updated data from our COVID-19 Symptom Survey, and would be happy to walk through this data with our research director, if helpful.

Thanks again--and please let me know if there’s anything I’m missing or can follow up to clarify.
Understood. I thought we were doing a better job through responding to this — and we are working to get the data that will more clearly show the universe of the Covid content that’s highest in distribution with a clear picture of what percentage of that content is vaccine hesitancy content, and how we are addressing it. I know Andy told me that would take a bit of time to nail down and we are working on that universe of data. I will make sure we’re more clearly responding to your questions below.

Thanks for this. [EXTERNAL] Should be trying to land a time.

Will say I’m really mostly interested in what effects the interventions and products you’ve tested have had on increasing vaccine interest within hesitant communities, and which ones have shown promise. Really couldn’t care less about products unless they’re having measurable impact. And while the product safari has been interesting, at the end of the day, I care mostly about what actions and changes you’re making to ensure you’re not making our country’s vaccine hesitancy problem worse. I definitely have what I believe to be a non-comprehensive list of products you’re building but I still don’t have a good, empirical answer on how effective you’ve been at reducing the spread of vaccine-skeptical content and misinformation to vaccine fence sitters in the now-folded “lockdown.” If [EXTERNAL] can speak to those things, great. [EXTERNAL] hasn’t been able to, but I’m sure someone there can.

In the electoral context, you tested and deployed an algorithmic shift that promoted quality news and information about the election. This was reported in the New York Times and also readily apparent to anyone with cursory social listening tools. You only did this, however, after an election that you helped increase skepticism in, and an insurrection which was plotted, in large part, on your platform. And then you turned it back off. I want some assurances, based in data, that you are not doing the same thing again here.

Hi Rob,

Wanted to follow up on your additional questions about WhatsApp — responses to your questions embedded in line and in blue below, along with a few attachments that are discussed in-line. Happy to discuss further.

Also — happy to schedule our next session with [EXTERNAL] for Monday if you’re interested. I know she was hoping to bring her colleague [EXTERNAL] to brainstorm on some ideas with you and Courtney. We can do this Monday or anytime next week.
Thanks,

We also wanted to follow up on your questions about WhatsApp. I’m sure you’re already attuned to this, but think it’s worth noting some of the key differences between a private messaging app like WhatsApp, and social media like Facebook and Instagram. Approximately 90 percent of the messages sent on WhatsApp are one-to-one, and the majority of group chats include fewer than ten people. WhatsApp does not promote content, and users do not build audiences or discover new people as they would on social media.

Very aware 😊

You’re right that without being able to see the content of messages on WhatsApp, we’re not able to measure prevalence (and, relatedly, reduction) of particular types of content. WhatsApp seeks to control the spread of misinformation and inform users through deliberate, content-agnostic product interventions -- things like labeling and limiting message forwards. The underlying idea there is that messages that did not originate from a close contact are less personal compared to typical messages sent on WhatsApp, and may be more prone to contain misinformation. The labels (“forwarded”; and “forwarded many times” if the message has been forwarded five times or more) are intended to prompt people to stop and think when they are reading a message and before they forward something, which may not be accurate. The forward limits (no more than five chats at a time; one chat a time for highly forwarded messages), are intended to reduce their spread. As mentioned in my earlier note, when WhatsApp rolled out the limitation for highly forwarded messages to one chat at a time in April 2020, this resulted in a 70% reduction of those messages globally. Of course, not all forwards are misinformation, so these are by nature somewhat blunt tools, but they are important ones -- and ones that many other messaging services don’t provide.

A few additional things to note:

1. WhatsApp also employs best-in-class spam detection technology to spot accounts engaging in mass messaging behavior, so they can’t be used to spread spam or viral misinformation. We ban over 2 million accounts per month for bulk messaging behavior, 75% of them without a recent user report, which means our automated systems stop abuse before users can report them. (This white paper describes these systems in further detail.)

We have a thing where we can’t click links from emails — can you send me the white paper?
White Paper is attached in PDF to this email.

2. Another aspect of what WhatsApp does -- again without accessing the content of messages -- is to provide tools to empower users to seek out reliable sources of information. One way we’ve done this in the product is through a “search the web” feature we rolled out last August, which allows users to easily double check highly forwarded messages they receive on WhatsApp by tapping a magnifying glass button in the chat to initiate a web search on their device browser. This helps users find news results or other sources of authoritative information about messages they have received from outside their close contacts -- and is available in English, Spanish, and other languages.
Can you show me what this might look like? What kind of testing have you seen around effectiveness? Are there other tactics you’ve deployed? Does exposure to forwarded messages change in any way the kinds of positive information they’re exposed to on Facebook or Instagram?

Attached is an image explaining how “Search the Web” functions on WhatsApp - and you can find more info at this link: >>https://blog.whatsapp.com/search-the-web/?lang=en<<. As we have rolled out Search the Web over the past year, we have conducted research - through interviews and surveys - to understand how users interact with this feature, what level of awareness they have about it and particularly, how it is used by low digital literacy users. Along similar lines, we are continuing to experiment with different forward depths that classify a message as a “Highly Forwarded Message” and bring up the magnifying glass button for that message. We will use these insights to design further product features that limit virality on WhatsApp.

With respect to your question about COVID-related information people may be exposed to Facebook and Instagram, that is not related to users’ personal messaging activity on WhatsApp.

3. WhatsApp also has partnerships with fact checking organizations, government agencies, and international organizations, like the WHO, around the world to make authoritative information about COVID-19 and vaccines available via WhatsApp. WhatsApp donated $1M to the International Fact Checking Network (IFCN) to support the CoronaVirusFacts Alliance, which brought together more than 100 fact checkers in 70+ countries in 40+ languages. These organizations have produced 9,000+ unique fact checks, all of which are accessible through a global fact-checking bot jointly created by the IFCN and WhatsApp.

How do they make the information available?

COVID-19 information is made available on WhatsApp by WHO, government health ministries, and third-party fact checkers through our WhatsApp Business API solution, which supports two-way conversational messaging and one-way notifications. These organizations access our API through approved business solutions providers (BSPs) to build chatbots on the WhatsApp Business API that are capable of returning automated responses to user queries. We support government partners by waiving WhatsApp fees associated with the API and making available Facebook ads credits to publicize these chatbots. For some fact checkers, we cover the BSP and end client costs through annual grants.

Users click on a link on the organization’s website to open the chat or text “hi” to the chatbot’s phone number. This brings them to a greeting message where they are presented with options to search for information on a COVID-related topic, access latest fact checks, or get tips to fight misinformation, among other things. The requested information is then provided in a variety of ways.

The WHO Health Alert on WhatsApp, for example, provides information about how vaccines work and how they are tested as a text message in response to a user query. It also provides users with links to videos of WHO’s “Science in 5” series where scientists discuss commonly asked questions about the Covid-19 Vaccines. The latest edition of this discussion is also sent to the user’s chat as an audio clip for ease of access.

The IFCN chatbot which leverages the CoronaVirusFacts Alliance database of COVID-19 misinformation allows users to search for fact checks based on keywords and will provide the latest fact-checks from networks in the user’s country as determined by the user’s phone number.

Screenshots of the WHO Health Alert and IFCN chatbot are attached.
4. We’re very cognizant of WhatsApp’s use among immigrant communities in the U.S. and we’re focused on ensuring these sorts of resources noted above are available in Spanish as well as English. During the 2020 election we partnered with Univision and Telemundo to make IFCN’s election-related fact checks available in Spanish. Both Univision and Telemundo are now in the process of getting approved as certified IFCN fact checkers, which will enable them to set up their own Spanish-language fact checks directly on WhatsApp with financial support from Facebook. This will add to existing Spanish-language resources available via WhatsApp, including the search the web feature and the CoronaVirusFacts Alliance bot mentioned above.

Is this true in other languages? I’m thinking specifically about languages that have prevalence in south Asian countries. And in the electoral context, what did you do there that worked and you’re taking into this body of work?

We encourage our partners to make their resources available as widely as possible. The IFCN CoronaVirusFacts Alliance chatbot is already available in the US in 4 languages - English, Hindi, Spanish and Portuguese. The Search the Web feature is currently available in English, Spanish, German, Italian and French; we have been working to expand the feature and it’s available to South Asian language markets in Android Beta (~25M users) but the quality of search results is not yet high enough for a full launch.

US 2020 was the biggest fact checking effort that WhatsApp supported and we’re pleased that these efforts have helped to spur progress in the broader fact checking ecosystem. The partnerships we built with Telemundo and Univision, helped lead to both companies establishing their own specialized Spanish-language fact checking units - EL Detector and T Verifica, respectively - and hiring data analysts and translators to aid their fact checking efforts.

We are also proud of the work that we did with IFCN during the US 2020 election to help create a consortium of fact checkers, which allowed these organizations to pool resources and scale their operations. We have been building on the success of this model elsewhere in the world - including in India where we have worked with six Indian fact checking organizations to build a similar coalition that will consolidate fact checks and trends on a common website.

One other initiative we are focused on are partnerships with governments, private healthcare providers, and pharmacies to support COVID-19 vaccination efforts through chat tools on WhatsApp. We’ve launched these successfully so far in Indonesia, Brazil, South Africa, and Argentina, among other countries, and are very interested in exploring ways to replicate some of these efforts in the U.S., especially in boosting the vaccination effort within the Latinx community. We are in discussions with the CDC and with officials in California, Delaware, and Los Angeles, and we are keen to work together to expand the scope and reach of these partnerships.

I guess I have the same question here as I do on Facebook on Instagram. Do you guys think you have this under control? You’re obviously going to say yes to that, so I guess the real question is, as ever: how are you measuring success? Reduction in forwarding? Measured impact across Facebook properties?

On WhatsApp, reduction in forwards is just one of the signals that we use to measure how well we are doing in reducing viral activity on our platform. We also ban accounts that engage in mass marketing or scam behaviors - including those that seek to exploit COVID-19 misinformation. Our efforts in this space are more comprehensive than anything that our peers in private messaging or SMS do, and we are constantly innovating to stay ahead of future challenges.
We also track engagement with some of the tools available on WhatsApp that provide access to fact checks and other authoritative sources of information. For instance, 3 billion messages related to COVID-19 have been sent by governments, nonprofits and international organizations to citizens through official WhatsApp chatbots, and over 300 million messages have been sent over COVID-19 vaccine helplines on WhatsApp during the 1st quarter of 2021.
From: [REDACTED]@fb.com
Sent: 4/14/2021 5:23:05 PM
To: Flaherty, Rob EOP/WHO [REDACTED]@who.eop.gov
CC: Slavitt, Andrew M. EOP/WHO [REDACTED]@who.eop.gov
Subject: Re: [EXTERNAL] tucker

Thanks—I saw the same thing when we hung up. Running this down now.

Get Outlook for iOS

From: Flaherty, Rob EOP/WHO [REDACTED]@who.eop.gov
Sent: Wednesday, April 14, 2021 1:10:41 PM
To: [REDACTED]@fb.com
Cc: Slavitt, Andrew M. EOP/WHO [REDACTED]@who.eop.gov
Subject: tucker

Since we’ve been on the phone—the top post about vaccines today is Tucker Carlson saying they don’t work. Yesterday was Tomi Lahren saying she won’t take one. This is exactly why I want to know what “Reduction” actually looks like—if “reduction” means “pumping our most vaccine hesitant audience with Tucker Carlson saying it doesn’t work” then... I’m not sure it’s reduction!

Rob Flaherty
Director of Digital Strategy
The White House
Cell: [REDACTED]
Hey—I’m really sorry, I missed this ahead of the 11:00. We will definitely prioritize for future. And working on both immediate follow ups—running down question on Tucker and working on getting you report by end of week.

Get Outlook for iOS

-----

Given the briefing at 11 and Andy’s interest in joining, I am wondering if it might be good to consider pushing back. If we were to do that, would anything between noon and 1:30 work? If not, we can proceed and folks can join as they get free.

Great thanks—Courtney we will follow up on anything that comes out of the 11:00.

We have our press briefing this morning at 11 so I won’t be there.

Thanks for sending the stuff below. I just pinged CDC on the FAQ and we will share as soon as they have

I will be there, yes.
Just confirming with you both that 11:00 this morning still works? You should have calendar invites—Courtney I saw you were not on our invite but added you.

Hi Rob, Courtney,

Thanks for this quick response—it was super helpful in informing our overall strategy today. I have some responses in blue below. I’m looking forward to the meeting tomorrow and hoping we can spend some time responding to Rob’s feedback from last week as well as further discussing the J&J news and how we can hopefully partner together.

Courtney—as we discussed, we also wanted to send over some examples of content we see on our platform that we remove (misinformation & harm) as well as content we take other actions on, but do not remove (vaccine hesitancy). I have included some examples at the bottom of this email and happy to set up time to talk through this more with you as well, if helpful.

Talk soon,

---

Some kind of thing that puts the news in context if folks have seen it (like your current “COVID news” panel) that has 3-4 pieces of info (eg: Adverse events are very rare—6 cases out of nearly 7 million, the FDA and CDC are reviewing so it's health care providers know how to treat any of the rare events, this does not affect Pfizer or Moderna, which vaccinate via a different mechanism). Happy to provide what those things should be. If the ultimate product pulls in social from others, we’re happy to put something together there as well.

Thanks very much for the suggestion—we are consistently updating the news module to provide timely and relevant context to users, such as article(s) that provide context on the rarity of experiencing blood clots. We would love any suggestions you all would have on trends you’re seeing.

- CDC is working through an FAQ that we’d love to have amplified in whatever way possible—maybe through the COVID info panel.

Thanks—we’ll be on the lookout for the FAQ and can discuss tomorrow.

- A commitment from you guys to make sure that a favorable review reaches as many people as the pause, either through hard product interventions or algorithmic amplification

Would love to talk through this one a bit more. Our goal is to ensure that people have access to authoritative info about the vaccine. We’re looking forward to talking more tomorrow about our approach to sharing authoritative info and what we’ve done today in support of that goal given the J&J announcement.

More broadly: we share concern about knock-on effects and are curious to get a read from your CMU data about what you’re seeing and with whom. Moreover, I want to make sure you have eyes on what might be spinning off the back end of this—that the news about J&J doesn’t spin off misinformation. Would be great to get a 24 hour report-back on what behavior you’re seeing.

We will look to get you insights as soon as we have them. We are going to be watching to see how this plays out over the next couple of days. is joining tomorrow and plans to share a couple things we are seeing emerge from the
CMU survey and what we are going to be watching over the next few days. Also, we are proactively monitoring and seeing what themes emerge from content on-platform and happy to share out when we have stuff collected.

**VACCINE HESITANCY EXAMPLES:**

The following examples of content are those that do not violate our Misinformation and Harm policy, but may contribute to vaccine hesitancy or present a barrier to vaccination. This includes, for example, content that contains sensational or alarmist vaccine misrepresentation, disparaging others based on the choice to or to not vaccinate, true but shocking claims or personal anecdotes, or discussing the choice to vaccinate in terms of personal and civil liberties or concerns related to mistrust in institutions or individuals. We utilize a spectrum of levers for this kind of content that is both proportionate and also helps our users make informed decisions. Actions may include reducing the posts’ distribution, not suggesting the posts to users, limiting their discoverability in Search, and applying Inform Labels and/or reshare friction to the posts. Depending on the category of content, we scale our interventions to have the highest public health impact, while understanding that healthy debate and expression is important.
**Examples of Content Removed for Violating our Misinformation & Harm Policy**

The following are examples of posts we have removed for violation of our Misinformation & Harm Policy.
Scientists Warn of Potential COVID Vaccine-Related ‘Ticking Time Bomb’ • Children's Health Defense
Scientists Warn of Potential COVID Vaccine-Related 'Ticking Time Bomb' • Children's Health Defense
Thanks for reaching out. Andy might reply separately, but here’s some thoughts.

I’m putting our public messaging below, which will be updated and we’ll be sure to send to you.

But generally, I think some combo of the following would be helpful:

- Some kind of thing that puts the news in context if folks have seen it (like your current “COVID news” panel) that has 3-4 pieces of info (eg: Adverse events are very rare – 6 cases out of nearly 7 million, the FDA and CDC are reviewing so health care providers know how to treat any of the rare events, this does not affect pfizer or moderna, which vaccinate via a different mechanism). Happy to provide what those things should be. If the ultimate product pulls in social from others, we’re happy to put something together there as well.
As of April 12, nearly 7 million J&J doses have been administered. CDC and FDA are investigating 6 cases of an extremely rare type of blood clot in individuals after receiving the J&J vaccine. As CDC and FDA noted in their statement, right now these adverse events appear to be extremely rare. Out of an abundance of caution as they review these rare cases, CDC and FDA are recommending vaccine providers pause on administering the J&J vaccine. As FDA noted this morning, they hope to review this quickly over the next few days. This pause is important so health care providers know how to treat any individuals who may experience these rare events.

This announcement will not have a significant impact on our vaccination plan: J&J vaccine makes up less than 5 percent of the recorded shots in arms in the United States to date. Based on actions taken by the President earlier this year, the U.S. has secured enough Pfizer and Moderna doses for 300 million Americans. You can read the full statement from White House COVID-19 Response Coordinator Jeff Zients on the impact on supply here.

We will be back in touch soon to share additional resources and messaging on this issue, as well as our broader efforts to advance vaccine confidence and protect America’s health.

---

From: [redacted]@fb.com
Sent: Tuesday, April 13, 2021 12:21 PM
To: Rowe, Courtney M. EOP/WHO; [redacted]@who.eop.gov; Flaherty, Rob EOP/WHO
Subject: FW: [EXTERNAL] Re: Connecting

Courtney and Rob—making sure you also receive this message—we want to get ahead of this but also want to make sure we are amplifying the right messages. Let us know if helpful to connect quickly today?

From: [redacted]@fb.com
Date: Tuesday, April 13, 2021 at 12:18 PM
To: Slavitt, Andrew M. EOP/WHO
Cc: [redacted]@fb.com
Subject: Re: [EXTERNAL] Re: Connecting

Hi Andy

Hope this finds you well?
Re the J+J news, we're keen to amplify any messaging you want us to project about what this means for people - it obviously has the risk of exacerbating vaccine hesitancy, so we're keen to get ahead of the knock-on effect. Don't hesitate to tell me - or via your teams - how we can help to provide clarity/reassurance via Facebook.

All very best
Hey Rob—understood and sorry for the delay. The team has been heads-down since our conversation to produce the report we discussed on Wednesday afternoon. We are aiming to get you something tonight ahead of the weekend. We want to respond to your questions below as well but I have been hoping to get this work completed and then to schedule a call to discuss. Would that work?

These questions weren’t rhetorical

And sorry — if this was not one of the most popular posts about the vaccine on Facebook today, then what good is Crowdtangle?

said that Tom’s video was the most popular yesterday based on your data, which reflected what CT was showing. Tucker’s video was top on CT today. What is different about this video, then?

Sent from my iPhone

On Apr 14, 2021, at 11:29 PM, Flaherty, Rob EOP/WHO wrote:

I guess this is a good example of your rules in practice then — and a chance to dive in on questions as they’re applied.

How was this not violative? The second half of the segment is raising conspiracy theories about the government hiding that all vaccines aren’t effective. It’s not about just J&J. What exactly is the rule for removal vs demoting?

Moreover: you say reduced and demoted. What does that mean? There’s 40,000 shares on the video. Who is seeing it now? How many? How effective is that?

And we’ve gone a million rounds on this in other contexts so pardon what may seem like deja vu — but on what basis is “visit the covid-19 information center for vaccine resources” the best thing to tag to a video that says the vaccine doesn’t work?
Not for nothing but last time we did this dance, it ended in an insurrection.

Sent from my iPhone

On Apr 14, 2021, at 11:11 PM, [redacted]@fb.com wrote:

Making sure you receive--

From: [redacted]@fb.com>
Date: Wednesday, April 14, 2021 at 10:51 PM
To: Slavitt, Andrew M. EOP/WHO [redacted]@who.eop.gov>
Cc: [redacted]@fb.com>
Subject: Re: Tucker Carlson anti-vax message.

Hi Andy - have looked into this some more.

I realize it may be of limited comfort at this moment, but this was not the most popular post about vaccines on Facebook today. Our data is slightly lagging, and we’ll get back to you with more detail on this specific post tomorrow. Right now, it appears that it probably was among the top 100 most-viewed vaccine posts. I’m including a few examples of posts that were more popular today at the end of this note.

Regardless of popularity, the Tucker Carlson video does not qualify for removal under our policies. Following the government’s decision yesterday, we are allowing claims that the Johnson and Johnson vaccine causes blood clots, but we still do not allow categorical claims that it or other vaccines are unsafe or ineffective.

That said, the video is being labeled with a pointer to authoritative COVID information, it’s not being recommended to people, and it is being demoted.

The team is working on the follow ups from the meeting this morning, including more details on most viewed/ranked content on Facebook and [redacted] will be in touch shortly on that - I’m very keen that we follow up as we’d agreed, and I can assure you the teams here are on it.

Given the timeline that was provided today for further decision about the J&J vaccine, it would be great to get your guidance about what affirmative messages we should amplify right now. Consistent with the message we heard at the press conferences, we’re currently emphasizing the safety and efficacy of the Moderna and Pfizer vaccines in the Covid Information Center.

Popular Vaccine-Related Content on Facebook Today:

ABC: >>https://www.facebook.com/10160902498218312<<
NBC: >>https://www.nbcnews.com/healthy/health-news/what-do-if-you-got-johnson-johnson-vaccine-n1263927<<
CDC: >>https://www.facebook.com/10159031890151026<<
CBS: >>https://www.facebook.com/10159467409732010<<
Heather Cox Richardson: >>https://www.facebook.com/297363371758902<<

All v best
On 4/14/21, 10:52 AM, [Redacted]@fb.com wrote:

Ok - sorry to hear about call today, will dig in now.

On 4/14/21, 10:01 AM, "Slavitt, Andrew M. EOP/WHO" [Redacted]@who.eop.gov wrote:

Number one on Facebook. Sigh.

Big reveal call with FB and WH today. No progress since we spoke. Sigh.

Sent from my iPhone
Rob—thanks for catching up earlier and sorry for the delay in getting these back to you. We can schedule time to discuss any of this further if helpful.

How was the Tucker post not violative?
- while we remove content that explicitly directs people not to get the vaccine, as well as content that contains explicit misrepresentations about vaccines, we reviewed this content in detail and it does not violate those policies.

Moreover: you say reduced and demoted. What does that mean? There’s 40,000 shares on the video. Who is seeing it now? How many? How effective is that?
- The video received 50% demotion for seven days while in the queue to be fact checked, and will continue to be demoted even though it was not ultimately fact checked.

Why does CT tell a different story than our internal number?
- Crowdtangle shows engagement not views, and a simple text search for “vaccine” in Crowdtangle doesn’t have the same recall as our classifiers, i.e., doesn’t include all of the posts about vaccines. The data that we provided doesn’t include the Tucker Carlson video because our data pipelines don’t populate that quickly—we provided data for the week before. (The delay in data doesn’t mean we aren’t able to find and remove violating content in real time—our systems do this automatically).

Why label this content with a generic “visit the covid information center” message?
- Our more granular label about vaccine safety previously said “COVID-19 vaccines go through many tests for safety and effectiveness before they’re approved”. In light of the decision to pause the J&J vaccine, vaccine safety discussion evolved past “approval,” and we were concerned that this was a confusing/irrelevant message to be applying to content discussion the decision to pause J&J without revoking approval. We temporarily reverted to a more generic message and are updating the more specific label for posts about vaccine safety to say “COVID-19 vaccines go through many tests for safety and effectiveness and then are monitored closely” to try to adapt to the changing factual situation and evolving discussion. This new message is being rolled out and should appear instead of the generic label now.

These questions weren’t rhetorical

And sorry — if this was not one of the most popular posts about the vaccine on Facebook today, then what good is Crowdtangle?
said that Tomis video was the most popular yesterday based on your data, which reflected what CT was showing. Tucker's video was top on CT today. What is different about this video, then?

Sent from my iPhone

On Apr 14, 2021, at 11:29 PM, Flaherty, Rob EOP/WHO wrote:

I guess this is a good example of your rules in practice then — and a chance to dive in on questions as they’re applied.

How was this not violative? The second half of the segment is raising conspiracy theories about the government hiding that all vaccines aren’t effective. It’s not about just J&J. What exactly is the rule for removal vs demoting?

Moreover: you say reduced and demoted. What does that mean? There’s 40,000 shares on the video. Who is seeing it now? How many? How effective is that?

And we’ve gone a million rounds on this in other contexts so pardon what may seem like deja vu — but on what basis is “visit the covid-19 information center for vaccine resources” the best thing to tag to a video that says the vaccine doesn’t work?

Not for nothing but last time we did this dance, it ended in an insurrection.

Sent from my iPhone

On Apr 14, 2021, at 11:11 PM, wrote:

Making sure you receive--

From: @fb.com>
Date: Wednesday, April 14, 2021 at 10:51 PM
To: Slavitt, Andrew M. EOP/WHO
Cc: @fb.com>
Subject: Re: Tucker Carlson anti-vax message.

Hi Andy - have looked into this some more.

I realize it may be of limited comfort at this moment, but this was not the most popular post about vaccines on Facebook today. Our data is slightly lagging, and we’ll get back to you with more detail on this specific post tomorrow.
Right now, it appears that it probably was among the top 100 most-viewed vaccine posts. I’m including a few examples of posts that were more popular today at the end of this note.

Regardless of popularity, the Tucker Carlson video does not qualify for removal under our policies. Following the government’s decision yesterday, we are allowing claims that the Johnson and Johnson vaccine causes blood clots, but we still do not allow categorical claims that it or other vaccines are unsafe or ineffective.

That said, the video is being labeled with a pointer to authoritative COVID information, it’s not being recommended to people, and it is being demoted.
The team is working on the follow-ups from the meeting this morning, including more details on most viewed/ranked content on Facebook and will be in touch shortly on that. I'm very keen that we follow up as we'd agreed, and I can assure you the teams here are on it.

Given the timeline that was provided today for further decision about the J&J vaccine, it would be great to get your guidance about what affirmative messages we should amplify right now. Consistent with the message we heard at the press conferences, we're currently emphasizing the safety and efficacy of the Moderna and Pfizer vaccines in the Covid Information Center.

Popular Vaccine-Related Content on Facebook Today:

ABC: https://www.facebook.com/101609024982188122
NBC: https://www.nbcnews.com/health/health-news/what-do-if-you-got-johnson-johnson-vaccine-n1263927
CDC: https://www.facebook.com/10159031890151026
CBS: https://www.facebook.com/10159467409732010
Heather Cox Richardson: https://www.facebook.com/297363371758902

All the best

---

On 4/14/21, 10:52 AM, slavitt@fb.com > wrote:

Ok - sorry to hear about call today, will dig in now.

On 4/14/21, 10:01 AM, "Slavitt, AndrewM. EOP/WHO" wrote:

Number one on Facebook. Sigh.

Big reveal call with FB and WH today. No progress since we spoke. Sigh.

Sent from my iPhone
All – Thanks again for the conversation today.

We’ll look out for the top trends that you’ve seen in terms of misinformation around the vaccine.

To recap: As we move away from a supply problem toward a demand problem, we remain concerned that Youtube is “funneling” people into hesitance and intensifying people’s hesitancy. We certainly recognize that removing content that is unfavorable to the cause of increasing vaccine adoption is not a realistic—or even good—solution. But we want to be sure that you have a handle on vaccine hesitancy generally and are working toward making the problem better. This is a concern that is shared at the highest (and I mean highest) levels of the WH, so we’d like to continue a good-faith dialogue about what is going on under the hood here. I’m the on the hook for reporting out.

Just before we were meeting, this article from Buzzfeed popped, highlighting the Youtube misinformation that is spreading through the Vietnamese community. I think this brings up a question that I had in our first meeting about your capabilities around misinformation in non-English-speaking communities. Clearly, more work to be done here. Would love to get some insights from you on how you are tackling this problem across all languages—how your enforcement has differed in languages and what your roadmap to improvement is.

A couple of other things it would be good to have from you all:

- As mentioned up top, the top trends that you’re seeing in terms of misinformation/hesitance inducing content (Stanford has mentioned that it’s recently Vaccine Passports and J&J pause-related stuff, but I’m not sure if that reflects what you’re seeing)
- A deeper dive on reduction and its effectiveness. It’s helpful that you mentioned that watch time is your key metric. I believe you said you reduced watch time by 70% on “borderline” content, which is impressive. Obviously, the term “borderline” is moveable, but taking it for what it is: How does that track with vaccine-related content specifically (removing the “UFO stuff”). What has the comparative reduction in watch time on “borderline” vaccine topics been after your interventions? And what has the increase in watch time been on authoritative information?
- I appreciated your unequivocal response that you are not recommending anti-vaccine content and you are lifting authoritative information in both search and recommendations to all audiences. Related to the second bullet: to what extent have your ranking interventions been effective there? And, perhaps more critically, to what degree is content from people who have been given a “strike” still being recommended and shown in prominent search positions?
- I feel like I am not coming away with a very clear picture of how you’re measuring the effectiveness of uplifting authoritative information. I obviously buy the theory—but how did you arrive on info-panels as the best intervention? And to what extent are people clicking through after exposure to vaccine-hesitant content? What are you doing mechanically to boost the authoritative information? When you have relevant influencers speak to experts, I imagine (hope?) it’s not just putting the content out there and that you’re recommending it to people for whom it would be most relevant. How does that work?
- What are the general vectors by which people see the “borderline” content—or really just vaccine-skeptical content? Is it largely through recommendations? Search?

We are excited to continuing partnering with you on this work as we have via but we want to make sure that the work extends to the broader problem. Needless to say, in a couple of weeks when we’re having trouble
getting people to get vaccinated, we’ll be in the barrel together here. We’ve worked with a number of platform partners to track down similar information based on internal data, including partners of similar scale. I am feeling a bit like I don’t have a full sense of the picture here. We speak with other platforms on a semi-regular basis. We’d love to get in this habit with you. Perhaps bi-weekly?

Looking forward to more conversation.

-Rob

Rob Flaherty
Director of Digital Strategy
The White House
Cell: [redacted]
So I guess I have two questions here:

1. He references the “three” widest reach posts, of which I believe this is one:
   https://www.facebook.com/DeeBlock253/posts/3528944520539112

   For one, it’s still up and seems to have gotten pretty far. And it’s got 365k shares with four comments. We’ve talked about this in a different context, but how does something like that happen? The top post, the one from the Wisconsin news station, has 2.1 million comments. Am I looking at one instance of sharing (so, one of the 365,000 shares) or is this genuinely a post that has been shared nearly 400,000 times but only four people commented on it? What is your assessment of what is going on here?

   Won’t come as a shock to you that we’re particularly interested in your demotion efforts, which I don’t think we have a good handle on (and, based on the below, it doesn’t seem like you do either). Not to sound like a broken record, but how much content is being demoted, and how effective are you at mitigating reach, and how quickly? As I’ve said, I don’t think our position is that you should remove vaccine hesitant stuff. However, slowing it down seems reasonable. I just can’t describe what it means or how you know its working.

   Also, health groups: sure. But it seems more likely that anti-vax stuff is moving in groups that are not about health but are...mom centric, or other spaces. Strikes me as the issue here is less from single-use anti-vaccine accounts and more about people who...do other things and are also vaccine hesitant. Seems like your “dedicated vaccine hesitancy” policy isn’t stopping the disinfo dozen—they’re being deemed as not dedicated -- so it feels like that problem likely carries over to groups.

   As a last thing, I’d be interested in seeing this 100 ranking in terms of reach from things that you aren’t actively promoting in the info panel. EG: the unicef one’s reach is because you’re putting it in a big, giant box that says “Facebook” on it, versus the way it distributes naturally.

   Making sure you see this from [redacted] to Andy as well — around anytime to discuss any and all things...

   Hi Andy,

   Thanks to your team for sharing the research work with us - the team have spent some time reviewing these and I wanted to send over some details on where we’re developing work in this space (and where we aren’t).
Firstly, I know [REDACTED] has sent the latest version of the Top 100 content report to Rob yesterday evening and I wanted to send you a quick note on the three pieces of vaccine content that were seen by a high number of people before we demoted them. Although they don’t violate our community standards, we should have demoted them before they went viral and this has exposed gaps in our operational and technical process.

The teams have spent the last 24 hrs analysing these gaps and are making a number of changes starting next week, including setting up more dedicated monitoring for Covid vaccine content on the cusp of going viral, applying stronger demotions to a broader set of content, and setting up daily review and analysis so that we have a better real-time view of what is being seen by lots of people. I will be checking on this closely to make sure that these additional steps show results - the stronger demotions in particular should deliver real impact. Please let me know if you’d like to discuss any of this in more detail.

Returning to the points raised by the research - much of this is fair comment and actually includes many of the integrity efforts we’ve already deployed and are actively improving on, or are related to planned launches in the coming months.

**Non-English mis/disinformation circulating without moderation (Spanish, Arabic, Chinese, among others) and; ISD reports evidence of the global threat that anti-vaccination disinformation and misinformation represents across languages and borders:** Rolling our efforts out globally and in other countries will take us some time, given the complexity and scale - we think that this will take a number of months before we’ve fully scaled this work and we are prioritizing languages where we know vaccine hesitancy is likely to be higher based on external data.

**Do not distribute or amplify vaccine hesitancy, and Facebook should end group recommendations for groups with a history of COVID-19 or vaccine misinformation:** Much of the research you shared called on us to ensure that our systems don’t amplify vaccine hesitancy content and this is top of mind for us. In addition to the changes I mentioned above, we have already removed all health groups from our recommendation feature on Facebook, and on Instagram we filter vaccine-related accounts from our “accounts you may follow feature”. We also remove accounts that may discourage vaccination from search features. We currently enforce on hashtags we know are shared to promote vaccine hesitancy content and are working to improve our automated systems here.

**Monitoring events that host anti-vaccine and COVID disinformation:** From our analysis, events do not make up a high proportion of borderline vaccine content that people see on Facebook right now, but we are working to improve automatic detection for events hosting anti-vaccine and COVID content. Our viral monitoring efforts will also help us detect events that are gaining views on Facebook, and we do remove events coordinating in-person gatherings that involve or encourage people who have COVID-19 to join.

**12 accounts are responsible for 73% of vaccine misinformation:** Lastly, we continue to review accounts associated with the 12 individuals identified in the CCDH “Disinformation Dozen” report, but many of those either do not violate our policies or have ceased posting violating content. Our “Dedicated Vaccine Discouraging Entity” policy is designed to remove groups and pages that are dedicated to sharing vaccine discouraging content and we continue to review and enforce on these where we become aware of them.

I realise that our position on this continues to be a particular concern for you which is why our teams regularly engage with a range of experts to check whether we are striking the right balance here. In early March, for instance, we discussed our planned approach with members of the “High Level Panel on Vaccine Confidence & Misinformation” (organized by London School of Hygiene and Tropical Medicine and the Center for Strategic and International Studies) and we have checked more recently with [REDACTED] of the Vaccine Confidence Project too.

Among experts we have consulted, there is a general sense that deleting more expressions of vaccine hesitancy might be more counterproductive to the goal of vaccine uptake because it could prevent hesitant people from talking through their concerns and potentially reinforce the notion that there’s a cover-up (especially, though not exclusively, in the US). Given how complicated this continues to be, especially due to the recent news cycle about the safety of some vaccines, we will of course continue to speak with experts on our position here and adapt our approach as needed.
Hope this update is helpful—and obviously I’m happy to speak any time.

Best

On 4/27/21, 3:33 AM, "Slavitt, Andrew M. EOP/WHO" wrote:

Thanks I assume you may have staff there. I hope they are well.

Sent from my iPhone

> On Apr 27, 2021, at 12:11 AM, @fb.com wrote:
> > Hi Andy
> > > I know you’re focusing on India a fair amount. Just fyi, we’re doing the following:
> > > - Amplifying localized authoritative information and services specific to this crisis (e.g., symptom triage information/when to go or not go to a hospital given systems are overwhelmed) on platform and via ad credits;
> > > - Activating WhatsApp Bots for symptom tracking and to connect users to nearby health resources;
> > > - Curating relevant content across CIC, News, and Latest Updates for India;
> > > - Proactively reviewing misinformation content in English, Hindi, and Bengali; and
> > > - Making an up to $10M financial contribution to support some immediate needs in country (e.g., extending medical supplies to underprivileged, augmenting oxygen supply shortages, etc.)
> > > And is keen to see what more we can do
> > > https://www.facebook.com/zuck/posts/10112926954780791
> > > & team are in touch with USAID - but don’t hesitate to point us to other next steps where we could be helpful.
> > > We also received the recommendations/observations from the research organizations you met re covid misinfo etc this afternoon - the teams are now looking at them carefully, and I’ll get back to you once that’s done.
> > > Best
> > >
> > > On 4/22/21, 7:23 PM, "Slavitt, Andrew M. EOP/WHO" wrote:
> > > I will arrange a call. Please let know the information on who to include. Thanks
> > > Sent from my iPhone
> > >
> > > On Apr 22, 2021, at 7:58 PM, @fb.com wrote:
> > >
> > > Hi Andy
> > >
> > > As promised, more info from below and slides re the CZI work attached. Do tell me how an useful connection can be made.
> > >
Thx

Thanks for looking into this. CZI has been working in this area since before the pandemic. We built IDSeq (Link: https://www.discoveridseq.com/) to sequence unknown pathogens and then adapted it to do genomic sequencing for COVID and California Departments of public health. Right now we are working with local departments that are deploying these funds to build up their internal capacity. However, we can't figure out if there is a centralized vision of how all of these individual efforts are supposed to come back together and if they do what the public officer facing tool is. Slides on the issue we are trying to address is also attached.

Would love to try to learn about any central plan to ensure that our work ends up being compatible and share back any learning if helpful.
Sure. They’re first connected to authoritative information, but then you, as of last night, were presenting an anti-vaccine account with less than 1000 followers alongside, at level, with those pinned accounts!

Here’s the thing. You know and I know that the universe of undecided people searching Instagram for “vaccines” – as compared to, say, Google -- is probably low. But “removing bad information from search” is one of the easy, low-bar things you guys do to make people like me think you’re taking action. If you’re not getting that right, it raises even more questions about the higher bar stuff. You say in your note that you remove accounts that discourage vaccination from appearing in recommendations (even though you’re using “primarily” to give yourself wiggle room). You also said you don’t promote those accounts in search. Not sure what else there is to say.

Youtube, for their warts, has done pretty well at promoting authoritative info in search results while keeping the bad stuff off of those surfaces. Pinterest doesn’t even show you any results other than official information when you search for “vaccines.” I don’t know why you guys can’t figure this out.

From: Flaherty, Rob EOP/WHO
Sent: 5/12/2021 12:52:18 PM
To: Rowe, Courtney M. EOP/WHO
Cc: Rowe, Courtney M. EOP/WHO
Subject: RE: [EXTERNAL] FB Newsroom post tomorrow re: our Covid work

Thanks Rob – both of the accounts featured in the tweet have been removed from Instagram entirely for breaking our policies. We’re looking into what happened.

Taking a step back, when searching for terms related to vaccines on Instagram, people are first connected with resources from experts. That means that before anything, if someone is looking to get information about COVID-19 or vaccines, they are encouraged to seek that information out from the most credible sources. To do this, anyone who searches for information related to COVID-19 or vaccines on Instagram is first shown an educational pop-up on top of search results connecting them, in the U.S., to the CDC website (as shown in the tweet). We’ve also pinned authoritative accounts in the top search results which is why you also see the CDC and Gavi, the Vaccine Alliance Instagram accounts first in the results page.

We are continuing to develop technology to improve the quality of search results at scale across Instagram – this is a continual process built on new technology to address adversarial accounts. Our goal is to not recommend accounts like those shown in the tweet in search, which again shouldn’t have been on our platform to begin with. We also remove accounts that may discourage vaccination from search by developing and using this new technology to find accounts on Instagram that discourage vaccines, and remove these accounts from search altogether. We’ve also removed accounts that primarily discourage vaccination from appearing where we recommend new accounts to follow, such as accounts you may like, and suggested accounts.

We clearly still have work to do to, but wanted to ensure you were aware of the authoritative resources we’re pointing people to first as we continue investing in removing accounts from search that may discourage vaccination.
From: Flaherty, Rob EOP/WHO <[redacted]@who.eop.gov>
Date: Tuesday, May 11, 2021 at 8:08 PM
To: fb.com>
Cc: Rowe, Courtney M. EOP/WHO [redacted]@who.eop.gov>
Subject: Re: [EXTERNAL] FB Newsroom post tomorrow re: our Covid work

Hard to take any of this seriously when you’re actively promoting anti-vaccine pages in search

>https://twitter.com/jessreports/status/1392182161512361984?s=21<

Sent from my iPhone

On May 10, 2021, at 7:53 PM, [redacted]@fb.com> wrote:

Rob and Courtney — I wanted to preview a newsroom post and some additional press outreach that we plan to put out tomorrow with some updates on our Covid efforts - a large part of which will be focused on what we’ve been doing to help meet vaccination goals.

Since January, we and our partners have been using trusted messengers and personalized messaging on our platforms to increase vaccine acceptance, and we’re seeing positive impact at scale. For example:

- Over 3.3 million people have visited the vaccine finder tool since its launch on March 11, using it to get appointment information from a provider’s website, get directions to a provider, or call a provider. In addition, we’re showing people reliable information about whether and when they’re eligible to get vaccinated through News Feed promotions and our COVID-19 Information Center. West Virginia’s Department of Health and Human Resources reported that their vaccine registrations increased significantly after Facebook started running these notifications.
- Since January, we’ve provided more than $30 million in ad credits to help governments, NGOs and other organizations reach people with COVID-19 vaccine information and other important messages. These information campaigns resulted in an estimated 10 billion ad impressions globally.
- More than 5 million people globally have used these profile frames. And more than 50% of people in the US on Facebook have already seen someone use the COVID-19 vaccine profile frames. We spun up this effort in partnership with HHS/CDC after public health experts told us that people are more likely to get a vaccine when they see someone they trust doing it.
- As you know, since April 2020, we’ve been collaborating with Carnegie Mellon University and University of Maryland on a global survey of Facebook users to gather insights about COVID-19 symptoms, testing, vaccination rates and more. In the US:
  - Vaccine acceptance has been increasing steadily since January, increasing nearly 10% among all US adults.
  - We observed a particularly large increase in vaccine acceptance within certain populations in the US. Vaccine acceptance increased 26% among Black adults and 14% among Hispanic adults.
  - Vaccine access also remains a challenge. Among adults who intend to get vaccinated (but have not yet), 36% feel uninformed about how to get a vaccine and only 22% reported that they have an appointment in April.

We saw the announcement last week of the 70% goal, and we’re eager to help support your efforts to reach that goal by July 4th. In particular, through our work on both voter registration and vaccines, we’ve had success with a targeted strategy for our in-product messages. If there are specific states/regions (or other population segments) you’re targeting to reach that goal that you can share with us, we can look at how we might be able to adjust our in-product efforts to
help amplify your efforts. We’d be happy to schedule a follow-up call with the right people to drill down on how we might be able to help with these efforts.

As always let me know if you have any questions.

Thanks,
Thanks. Hi Anita and Rob — definitely agree and look forward to connecting.

Sending a post that went live this afternoon with information that I know we’ve discussed in the past. We had a conversation with the Surgeon General’s office yesterday to discuss the advisory in more detail and hope to continue to work to address concerns.

Along with — I am really hoping to close the gap in terms of what’s playing out publicly and what we might be able to accomplish working together.

Rob—I’m around anytime for a conversation.


---

Get Outlook for iOS

---

Thanks Anita, and thanks Rob. I appreciate the willingness to discuss. We’d love to find a way to get things back to a productive conversation. Adding in — to help us here — obviously Rob and have a tight working relationship already.

---

Hi — Happy to connect.

Sent from my iPhone

On Jul 17, 2021, at 5:56 PM, Dunn, Anita B. EOP/WHO wrote:

Hi — and thanks for reaching out. I’m adding Rob Flaherty, our Office of Digital Services Director, to this chain as well because he has been following your platform (and others) closely when it comes to flow of information and misinformation.
Perhaps it makes sense to schedule a conversation?

Anita

From: [redacted]@fb.com>
Sent: Saturday, July 17, 2021 5:52 PM
To: Dunn, Anita B. EOP/WHO [redacted]@who.eop.gov>
Subject: [EXTERNAL] hoping to connect

Hi Anita - hope you are well,

Would love to connect with you on the President’s comments on Covid misinformation and our work there. Really could use your advice and counsel on how we get back to a good place here.

While there’s always been a disagreement on where the lines should be on misinformation generally, we have genuinely tried to work with the administration in good faith to address the gaps and solve the problems. As I hope you know, we’ve been doing a significant amount of work to both fight the misinformation and fight the pandemic through authoritative information. Obviously, yesterday things were pretty heated, and I’d love to find a way to get back to pushing together on this - we are 100% on the same team here in fighting this and I could really use your advice.

Thanks,
Rob,

To clarify, the content was not in violation of our policies and therefore not subject to removal. But for all content on YouTube, we apply our 4R framework we have previously described to raise authoritative voices while reducing visibility on borderline content. External evaluators use these guidelines which are then used to inform our machine learning systems that limits the spread of borderline content.

Best Regards,

On Tue, Jul 20, 2021 at 8:36 PM Flaherty, Rob R. EOP/WHO <who.eop.gov> wrote:
So this actually gets at a good question — the content [redacted] points out isn’t defined as “borderline” and therefore isn’t subject to recommendation limitations?

Sent from my iPhone

On Jul 20, 2021, at 8:27 PM, [redacted] <google.com> wrote:

Rob -

I'll check with our team and share any additional data points we have available. Per our COVID-19 medical misinformation policy, we will remove any content that contradicts local health authorities’ or the World Health Organization’s (WHO) medical information about COVID-19. To date, approximately 89% of videos removed for violations of this policy were removed with 100 views or less. With regards to the specific videos you referenced, the content was not in violation of our community guidelines.

Best Regards,

On Tue, Jul 20, 2021 at 3:58 PM Flaherty, Rob R. EOP/WHO <who.eop.gov> wrote:
I see that’s your goal – what is the actual number right now?

I guess: does the content that [redacted] references in his tweet count as violative content that has slipped through? Or is it that generally the stuff he’s posting is in-bounds?
Thanks Rob,

We appreciate your interest in our announcement yesterday. With regards to your question on the Tweet, it is important to keep in mind that borderline content accounts for a fraction of 1% of what is watched on YouTube in the United States. We use machine learning to reduce the recommendations of this type of content, including potentially harmful misinformation. In January 2019, we announced changes to our recommendations systems to limit the spread of this type of content which resulted in a 70% drop in watchtime on non-subscribed recommended content in the U.S. and our goal is to have views of non-subscribed, recommended borderline content below 0.5%. I will keep you updated with any new policy or product improvements that we make as we continue our work to help people find authoritative health information on YouTube.

Best Regards

On Tue, Jul 20, 2021 at 10:57 AM Flaherty, Rob R. EOP/WHO wrote:

I'm curious: Saw this tweet. >>>https://twitter.com/ddale8/status/1417130268859772929<<<;

I think we had a pretty extensive back and forth about the degree to which you all are recommending anti-vaccination content. You were pretty emphatic that you are not. This seems to indicate that you are. What is going on here?

Thanks!

-Rob
We wanted to share an announcement that we recently made regarding a few new ways in which we are making it easier for people to find authoritative information on health topics on YouTube.

Starting this week, you’ll see two new features next to some health-related searches and videos. These include a new health source information panel that will surface on videos to provide context about authoritative sources, and a new health content shelf that more effectively highlights videos from these sources when you search for specific health topics. These context cues are intended to help people more easily navigate and evaluate credible health information.

To identify the sources that will be eligible to be included in these new features, we applied the principles recently developed and published by an expert panel convened by the National Academy of Medicine.

You can find more information about our announcement here. We’d be happy to set up time to walk you through these new features or answer any questions you may have - please let me know what works best for you.

Best Regards,

[Name]

---

[Name] Government Affairs & Public Policy Manager, YouTube [email protected]
Happy to.
Hoor, could you surface some times that work for your folks and we can go from there?

On Aug 2, 2021, at 6:04 PM, Tom, Christian L. EOP/WHO wrote:

Thanks or the info. A call might be helpful, if we can do something early next week? Adding Hoor here but appreciate your email and making time to talk further about it!

Hi All,

Per my and Christian’s phone call last Tuesday, I gathered more details for you and your team; happy to set up a call to discuss further as well.

As you know, we take aggressive steps to reduce the spread of vaccine hesitancy and vaccine misinformation on our platforms and we deploy technology to do so. As part of our efforts on Instagram, we have measures to help ensure we don’t recommend people follow accounts that promote vaccine hesitancy at scale. For two weeks in April (April 14-28) this measure was impacted by over-enforcement on a signal we used -- accounts that were posting far above normal vaccine-related content -- and removed these otherwise eligible accounts from being recommended as an account to follow. This did not impact reach or distribution of content in Feed or Stories or other areas of account discovery on Instagram, such as search or Explore.

Per your request for remediation, while we cannot boost your account in our recommendations, we are always here to help with content strategy, best practices, and further opportunities to collaborate.

Again, happy to discuss further on a call.

Best,
Hi Rob – I totally understand how frustrating that is. This was due to a bug in our recommendation surface, and was resolved in late May. Accounts affected did not specifically lose any followers as a result, nor was their presence reduced in Search or Explore, however. If you want to hop on the phone to discuss it, I’m at ___ anytime.

Hi Tegan – from what we understand it was an internal technical issue that we can’t get into, but it’s now resolved and should not happen again.

Thanks ___. Could you tell me more about the technical issues affecting audience growth? Was this just us and do you have a sense of what the issue was?
Hi again Tegan!

Coming back here on a few things:
- First, the technical issues that had been affecting follower growth on @potus have been resolved. Though there is still the issue of bot accounts being removed as normal, you should start to see your numbers trend back upwards, all things being equal and notwithstanding the big spike you saw this week given the collaboration with Olivia Rodrigo. Thanks for your patience as we investigated this.

- The answers to your aspect ratio, video quality and thumbnail questions can all be found in our Help Center here: https://www.facebook.com/help/instagram/381435875695118 and in the links on that page. Regarding 1:1 or 4:5 for feed video, I don’t have any specific recommendations on it. Obviously we know social managers are busy creating video for multiple platforms, so rest assured there is no algorithmic downside to using one crop over another.

- Finally, I can’t release any numbers related to the performance of different video formats, or light mode vs. dark mode usage unfortunately.

Let me know if you have any outstanding questions on these.

Appreciate it!

Hi Teagan! Let me round up some answers to these questions and come back to you shortly. Attached is the last edition of our IGTV video specs for you to check out in the interim.

Speak soon!

Hope you’re both well! I’m updating specs and guidelines for our video team and had a few quick questions.

- Do you have a guide/recommendation on codec/video quality? We’ve seen some issues with video files that display crisply on other platforms
- Do you have an updated thumbnail guide for IGTV and reels?
- Do you see any difference in performance between black, white, and branded video mattes on square videos in vertical placements?
- Do more people use night mode than day mode?
- For in-feed video (not sure what to call this but non-IGTV, non-reel video) do you recommend 1:1 or 4:5 these days?

Thank you!
Hi Rob -
I'm around if you'd like to dial me.

Best.

On Fri, Dec 17, 2021 at 5:33 PM Flaherty, Rob R EOP/WHO wrote:

New to the thread here, but this all reads to me like you all are bending over backwards to say that this isn’t causing confusion on public issues. If the AP deems it confusing enough to write a fact check, and you deem it confusing enough to create an event for it, how on earth is it not confusing enough for it to at least have a label?

Total Calvinball.

---

Thanks The policy at the top says:

What is in violation of this policy

In order for content with misleading media (including images, videos, audios, gifs, and URLs hosting relevant content) to be labeled or removed under this policy, it must:

- Include media that is significantly and deceptively altered, manipulated, or fabricated,
Include media that is shared in a deceptive manner or with false context, and

Include media likely to result in widespread confusion on public issues, impact public safety, or cause serious harm

I’ve highlighted the above sections which say that the first condition can be met alone OR the second and third can be met.

So that section that you’ve quoted makes sense, except this media is unto itself “significantly and deceptively altered, manipulated or fabricated.” And thus it should meet the criteria as outlined in the first bullet point.

Is that right?

Hi Christian,

I huddled with our enforcement teams on this who confirmed that the media does not meet our threshold for either significant or moderate risk of harm. Due to the low risk associated, the team found it to not meet the requirements for a label. They’ve specifically pointed to this language in our Help Center article:

"Tweets that share misleading media are subject to removal under this policy if they are likely to cause serious harm. Some specific harms we consider include:"

"Threats to physical safety of a person or group"
Incitement of abusive behavior to a person or group
Risk of mass violence or widespread civil unrest
Risk of impeding or complicating provision of public services, protection efforts, or emergency response
Threats to the privacy or to the ability of a person or group to freely express themselves or participate in civic events, such as:

Unfortunately, there isn't anything further here I can do in regards to our enforcement teams. If anything changes, we'll be sure to let you know. Appreciate your continued partnership and please don't hesitate to let us know if you have additional Tweets for review, anytime.

On Fri, Dec 17, 2021 at 3:49 PM Tom, Christian L. EOP/WHO wrote:
Hello!

Wanted to follow-up before we hit EOW. Even if this particular moment is not as much in the public eye right now, it's really important to us that this is addressed -- both on this particular one as well as a precedent for other moments when this might come up.

So, we appreciate your response and update here when you can provide.

Thanks,

--- Christian

From: [removed]@twitter.com>
Sent: Monday, December 13, 2021 4:05 PM
To: Tom, Christian L. EOP/WHO@who.eop.gov>
Cc: LaRosa, Michael J. EOP/WHO@who.eop.gov>
Subject: Re: [EXTERNAL] Re: Doctored video on Twitter of the First Lady

Hello! Apologies as I have been out of the office. I am working with the internal teams for clarity around your specific questions, so I will let you know as soon as I hear.

Appreciate your continued feedback here!
On Mon, Dec 13, 2021 at 12:16 PM Tom, Christian L. EOP/WHO wrote:

I hope you had a good weekend. Wanted to make sure we addressed this! Please let us know if you have a few mins to chat or if you can help us to make sure the enforcement of the policy is consistent.

From: Tom, Christian L. EOP/WHO
Sent: Thursday, December 9, 2021 4:37 PM
To: LaRosa, Michael J. EOP/WHO; Twitter.com
Subject: RE: [EXTERNAL] Re: Doctored video on Twitter of the First Lady

Hi

I wanted to follow-up here. Know this particular moment might have “passed” in terms of the scale/reach of it but in order to help us understand the Twitter processes best, would appreciate clarification on this when you’re able.

Thanks,

-- Christian

From: LaRosa, Michael J. EOP/WHO
Sent: Wednesday, December 1, 2021 5:13 PM
To: Twitter.com
Cc: Tom, Christian L. EOP/WHO
Subject: RE: [EXTERNAL] Re: Doctored video on Twitter of the First Lady

Thank you!
Michael LaRosa
The White House
Press Secretary | Office of the First Lady

Of course. Let me pass these additional questions along to the policy team directly for their insights and consideration. I’ll let you know from there!

On Wed, Dec 1, 2021 at 5:05 PM LaRosa, Michael J. EOP/WHO@who.eop.gov> wrote:

Thanks, Christian. Hi Let me know if we should hop on the phone to clarify. I am curious as to what would classify as “likely” so it is indisputable that the video is “deceptively altered,” “fabricated,” and “shared in a deceptive manner.”

Michael LaRosa
The White House
Press Secretary | Office of the First Lady
OK thanks - think this one does not fall under the “likely to impact public safety or cause serious harm” but it does fall under the first two in the chart, which includes “significantly and deceptively altered or fabricated” and “shared in a deceptive manner?”

And if the first two are met but the third is not, the chart says it is “likely to be removed.” Can you share any other info about why this one is not getting what Twitter would otherwise say is the “likely” outcome?

Also happy to chat on the phone this afternoon with Michael (who is the First Lady’s Press Secretary) if helpful.

Appreciate you following up. After escalating this to our team, the Tweet and video referenced will not be labeled under our synthetic and manipulated media policy. Although it has been significantly altered, the team has not found it to cause harm or impact public safety.

The team was able to create this Twitter Moment (here) and event page for more context and details:

>>>https://twitter.com/i/events/1435769009073123330<<<<;

Appreciate your feedback, as always.

On Wed, Dec 1, 2021 at 9:14 AM Tom, Christian L. EOP/WHO @who.eop.gov> wrote:

Just wanted to follow-up here.

It looks like from the rubric that this fits the first two criteria, which means it is “likely” to be labeled:
Thanks again

-- Christian

From: Tom, Christian L. EOP/WHO
Sent: Tuesday, November 30, 2021 8:54 PM
To: twitter.com; LaRosa, Michael J. EOP/WHO
Subject: RE: [EXTERNAL] Re: Doctored video on Twitter of the First Lady

Thanks Will you apply the “Manipulated Media” disclaimer to the video asset itself?

Both the linked tweet below and the original source of the video:

>>>https://twitter.com/PapiTrumpo/status/1465439569965424643<<<

Thanks

-- Christian

From: twitter.com
Sent: Tuesday, November 30, 2021 7:31 PM
To: LaRosa, Michael J. EOP/WHO
Cc: Tom, Christian L. EOP/WHO
Subject: Re: [EXTERNAL] Re: Doctored video on Twitter of the First Lady

Update for you - The team was able to create this event page for more context and details: >>>https://twitter.com/i/events/1465769009073123330<<<
On Tue, Nov 30, 2021 at 4:23 PM LaRosa, Michael J. EOP/WHO wrote:

Thank you!

Michael LaRosa
The White House
Press Secretary | Office of the First Lady
@who.eop.gov

From: twitter.com>
Sent: Tuesday, November 30, 2021 4:04 PM
To: Tom, Christian L. EOP/WHO @who.eop.gov>
Cc: LaRosa, Michael J. EOP/WHO @who.eop.gov>
Subject: [EXTERNAL] Re: Doctored video on Twitter of the First Lady

Hi Christian,

Happy to escalate with the team for further review from here.

Don't hesitate to let me know if you have any additional questions in the meantime.

On Tue, Nov 30, 2021 at 3:58 PM Tom, Christian L. EOP/WHO wrote:

Hi

Would you mind looking at this video and helping us with next steps to put a label or remove it?

https://twitter.com/ArtValley818/status/1465442266810486787?s=20
For reference, the timestamp is 32:47 for the undoctored video source here:


Thanks,

-- Christian
Happy to talk through it but if your product is appending misinformation to our tweets that seems like a pretty fundamental issue

On Aug 11, 2022, at 1:23 PM, @twitter.com wrote:

Hi Rob,

Thanks for reaching out. I believe you're referring to our Birdwatch product feature. Here's the latest information about how it works.

We'd be happy to arrange a meeting to walk you through how it works. We're also collecting feedback for our teams.

Best,
Hi Jesse- I just tried you on your cell. I'm at [redacted].

Best,

[redacted]

On Thu, Aug 11, 2022 at 1:28 PM Lee, Jesse C. EOP/WHO <[redacted]@who.eop.gov> wrote:

Thanks [redacted] I like the feature! But this note is factually inaccurate. This is a very technical question but you don’t have it right, and you are in effect calling the President a liar when his tweet is actually accurate. I’m happy to discuss this with whoever is the right person.

Cell: [redacted]

Sent from my iPhone

On Aug 11, 2022, at 1:23 PM, [redacted]@twitter.com wrote:

Hi Rob,

Thanks for reaching out. I believe you're referring to our Birdwatch product feature. Here's the latest information about how it works.

We'd be happy to arrange a meeting to walk you through how it works. We're also collecting feedback for our teams.

Best,

[redacted]

On Thu, Aug 11, 2022 at 12:31 PM Flaherty, Rob R. EOP/WHO <[redacted]@who.eop.gov> wrote:

Adding [redacted] since [redacted] seems to be out

> On Aug 11, 2022, at 12:21 PM, Flaherty, Rob R. EOP/WHO <[redacted]@who.eop.gov> wrote:
> Happy to connect you with some economists who can explain the basics to you guys
>
> https://mobile.twitter.com/gasbuddyguy/status/1555541573835886592/photo/1