

Taking action to combat COVID-19 vaccine misinformation

At Facebook, we're working to help people get vaccinated by improving access to information about vaccines and how to get vaccinated, making it easier for people to share their support for vaccines, and reducing misinformation about vaccines. While the Surgeon General said "*Limiting the spread of health misinformation is a moral and civic imperative that will require a whole-of-society effort*," we've already taken action on the [eight recommendations from the Surgeon General](#) about what technology platforms can do on COVID-19 misinformation.

1. **Assess the benefits and harms of products and platforms and take responsibility for addressing the harms.** In particular, make meaningful long-term investments to address misinformation, including product changes. Redesign recommendation algorithms to avoid amplifying misinformation, build in "frictions" -- such as suggestions and warnings -- to reduce the sharing of misinformation, and make it easier for users to report misinformation.

We are consistently updating our policies and systems to remove and/or reduce harmful content, including COVID misinformation, not to amplify it. We use a combination of artificial intelligence, human review, and input from partners -- including fact checkers and health organizations -- to address problematic content. We are constantly improving our systems. We review proposed product changes carefully to ensure they have a positive impact.

We have taken numerous steps to ensure people see reliable, high quality content about COVID-19 and how to get vaccinated - and those efforts are working:

- a. 3.3 million people have visited our vaccine finder tool to get appointment information from a provider's website, get directions to a provider or call a provider about COVID vaccines.
- b. 2 billion people have connected with reliable information about COVID and vaccines through our COVID Info Center
- c. More than 5 million people in the US, and more than 10 million people globally, have used COVID-19 vaccine profile frames to share their support for vaccines. And more than 50% of people in the US on Facebook have already seen someone use the COVID-19 vaccine profile frames, which we developed in collaboration with the US Department of Health and Human Services and Centers for Disease Control and Prevention.

In April 2020, we started showing messages in News Feed to people who liked, commented on or reacted to posts with misinformation about COVID-19 on Facebook that we removed for violating our policy.

We add labels to Facebook and Instagram posts that discuss the COVID-19 vaccines. These labels include reliable information about the safety of COVID-19 vaccines from the World Health Organization. For example, we're adding a label on posts that discuss the safety of COVID-19 vaccines that notes COVID-19 vaccines go through tests for safety and effectiveness before they're approved. We've also added an additional screen when someone goes to share a post on Facebook and Instagram with an informational COVID-19 vaccine label. It provides more information so people have the context they need to make informed decisions about what to share.

Each time a fact-checker rates a piece of COVID-19 content as false, Facebook significantly reduces the content's distribution so that fewer people see it. We notify people who previously shared the content or try to share it that the

information is false, and apply a warning label that links to the fact-checker’s article, disproving the claim with original reporting. We also use AI to scale the work of fact-checkers by applying warning labels to duplicates of false claims, and reducing their distribution.

Pages, groups, accounts, or websites that repeatedly share content rated False or Altered by fact-checkers will have some restrictions, including having their distribution reduced. Pages, groups, and websites may also have their ability to monetize and advertise removed, and their ability to register as a news Page removed.

We want to give people more information before they like a Page that has repeatedly shared content that fact-checkers have rated, so you’ll see a pop up if you go to like one of these Pages. You can also click to learn more, including that fact-checkers said some posts shared by this Page include false information and a link to more information about our fact-checking program. This will help people make an informed decision about whether they want to follow the Page.

Users can easily report misinformation on any piece of content by clicking on the three dots on the top right corner, clicking “Find support or report post,” then “False information.”

-
2. **Give researchers access to useful data to properly analyze the spread and impact of misinformation.** Researchers need data on what people see and hear, not just what they engage with, and what content is moderated (e.g., labeled, removed, downranked), including data on automated accounts that spread misinformation. To protect user privacy, data can be anonymized and provided with user consent.

CrowdTangle is a tool from Facebook that makes it easy for academics, journalists, and the public to follow, analyze, and report on what’s happening with public content on social media content. People can find larger trends to understand how public content spreads on social media. We have several COVID-19 Live Displays publicly available to see what content is being shared on social media about the virus.

Since April 2020, we’ve been collaborating with Carnegie Mellon University and University of Maryland on a global survey to gather insights about COVID-19 symptoms, testing, vaccination rates and more. This is the largest survey of its kind, with over 60 million total responses, and more than 170,000 responses daily across more than 200 countries and territories. This effort generates localized insights for researchers, public health officials and policymakers who are working to end the pandemic as quickly as possible.

-
3. **Strengthen the monitoring of misinformation.** Platforms should increase staffing of multilingual content moderation teams and improve the effectiveness of machine learning algorithms in languages other than English since non-English-language misinformation continues to proliferate. Platforms should also address misinformation in live streams, which are more difficult to moderate due to their temporary nature and use of audio and video.

We use experienced teams and technology to find and remove misinformation that breaks our rules on COVID and vaccine misinformation. While we’re constantly working to improve our systems, we believe we have some of the strongest systems in place to identify and take action against misinformation that breaks our rules.

In addition, we work with more than 80 global fact-checking partners to review and debunk misinformation, including claims about COVID and vaccines, in more than 60 languages. Since the beginning of the pandemic, we’ve labeled hundreds of millions of pieces of COVID-19 content rated false by our network of fact checking partners.

We’re operating our entire comprehensive strategy to combat COVID-19 misinformation in Spanish. This includes running the largest online vaccine information campaign in history and enforcing our policies and removing false claims about COVID-19 and vaccines - all in Spanish. We use the same machine learning model approaches in Spanish as we do

in English to remove misinformation that violates our Community Standards, and we have four US-based fact-checking partners who review and rate content in Spanish.

4. **Prioritize early detection of misinformation “super-spreaders” and repeat offenders.** Impose clear consequences for accounts that repeatedly violate platform policies.
-

We have strong policies to remove false claims about COVID and vaccines that we enforce - and when a Page, Group, or account repeatedly violates these policies we enforce penalties against them, including removing entities that repeatedly break our rules from our platforms.

Pages groups, accounts, or websites that repeatedly share content rated false by fact-checkers will have some restrictions, including having their distribution reduced. Pages, groups, and websites may also have their ability to monetize and advertise removed, and their ability to register as a news Page removed.

5. **Evaluate the effectiveness of internal policies and practices in addressing misinformation and be transparent with findings.** Publish standardized measures of how often users are exposed to misinformation and through what channels, what kinds of misinformation are most prevalent, and what share of misinformation is addressed in a timely manner. Communicate why certain content is flagged, removed, downranked, or left alone. Work to understand potential unintended consequences of content moderation, such as migration of users to less-moderated platforms.
-

We have clear policies in place for what violates our rules on COVID and vaccine misinformation that are publicly posted, including in our [Help Center](#), and we always aim to remove violating content as quickly as possible. We use a combination of people and technology to find and remove this violating content as quickly as possible and we’ve removed millions of pieces of content to date. We’ve also continued to adjust the list of claims that violate our policies to account for changing aspects of the pandemic. For example, we’ve recently added claims around theories that COVID vaccines make you ‘magnetic’ and viral ‘shedding’ that we remove from our apps.

We take a number of steps to identify, review, and remove potential misinfo before it goes viral, including temporary demotions of some content that is awaiting review by our third-party fact-checkers and monitoring the top viral content on our platform.

When COVID-19 misinformation does not present risk of contributing to imminent physical harm but is fact-checked by one of our partners, we label and reduce the visibility of that content. We also add informational labels to content about COVID-19 vaccines while promoting authoritative information on COVID-19 and vaccines. These measures play a key role in promoting understanding about COVID-19 and the COVID vaccine, further reducing the spread of misinformation.

6. **Proactively address information deficits.** An information deficit occurs when there is high public interest in a topic but limited quality information available. Provide information from trusted and credible sources to prevent misconceptions from taking hold.
-

An important component of our strategy to combat COVID-19 misinformation and to inform our community is to develop tools that promote vaccines and connect people to authoritative information from trusted sources:

Making it easier to get vaccinated: We’re providing authoritative information to help people find vaccine appointments in their area through News Feed messages. We’ve heard from states that this has had an impact: West Virginia reported their vaccine registrations increased significantly after we started promoting eligibility information in News Feed. The CDC launched a Spanish-language vaccine finder on WhatsApp, making it easy to find a location to get the shot, order a free ride and get information.

Helping people get questions answered: We've delivered 10B impressions worldwide of ads supporting partner campaigns since January. A single "Facts about COVID19" News Feed campaign increased belief in key facts about vaccine safety and testing by 3% across 5 countries. We've directed 2B+ people to expert health resources through the COVID Information Center.

Social norming through profile frames: We know from public health research that people are more likely to get vaccinated if they see others in their community doing so. Working with the HHS and CDC, earlier this year, we launched profile frames for people to share they've been or are planning to get vaccinated. In the US, more than 50 percent of Facebook users have seen someone they follow use one of these frames in News Feed. Since launch, 5M+ people in the US, including 10M+ people globally, have used these frames; vaccine stickers on IG have been used 7M+ times.

Supporting low vaccination rate communities: We're more frequently reaching people in areas with lower vaccination rates using CDC's Social Vulnerability Index. We are partnering on campaigns with KFF (Kaiser Family Foundation), featuring black doctors, nurses and researchers, and Spanish-language campaigns from Johns Hopkins University's Bloomberg School of Public Health.

Informing vaccination efforts through data: In collaboration with CMU and UMD, we're running the largest global health survey ever conducted, with over 60M total responses to date. This survey is used by Institute for Health Metrics and Evaluation (IHME) and governments around the world. In the US, broad trends are positive since January: vaccine acceptance has trended up overall and racial/ethnic disparities have reduced. We're also offering our Brand Lift measurement platform to help partners replicate and extend approaches that are working.

Trainings for people to spot misinformation: We're partnering with PEN America to offer media literacy training sessions for underserved communities around the US. Their Knowing the News program equips people with skills to evaluate the information they read about COVID-19 and vaccines.

-
7. **Amplify communications from trusted messengers and subject matter experts.** For example, work with health and medical professionals to reach target audiences. Direct users to a broader range of credible sources, including community organizations. It can be particularly helpful to connect people to local trusted leaders who provide accurate information.

Since January 2021, we've given 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. We're also adding authoritative information to posts about vaccines on Facebook and Instagram that link to the COVID-19 Information Center for more resources.

We're also providing authoritative information to help people find vaccine appointments in their area through News Feed messages. The CDC launched a Spanish-language vaccine finder on WhatsApp, making it easy to find a location to get the shot, order a free ride and get information.

-
8. **Prioritize protecting health professionals, journalists, and others from online harassment,** including harassment resulting from people believing in misinformation. communications from trusted messengers and subject matter experts.

Any journalist can register with Facebook to get important safety features that protect against harassment, hacking and other unusual account activity, and other issues journalists face online. More details [here](#). In April, we updated our policies to remove coordinated attempts to attack people for their choices around COVID vaccines as we believe this activity constitutes harassment and should be banned from our platforms.
