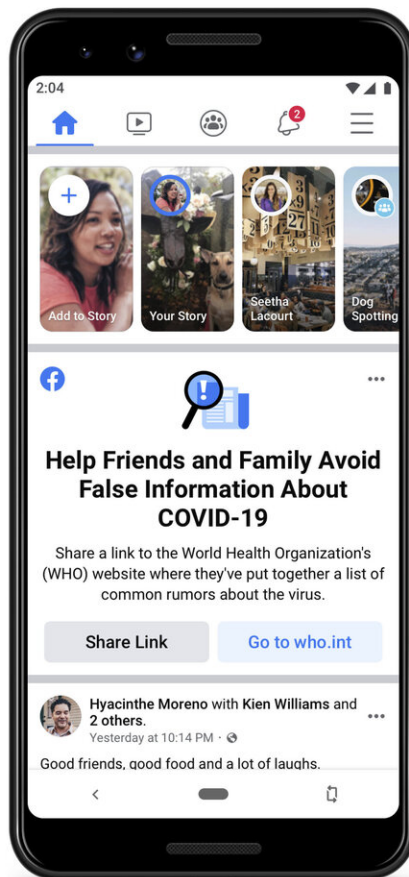# Facebook to warn users who 'liked' coronavirus hoaxes

April 16 2020, by Barbara Ortutay and Amanda Seitz

This undated image provided by Facebook shows the Facebook app on a cellular telephone. Facebook will soon let you know if you shared or interacted with dangerous coronavirus misinformation on the site, the latest in a string of aggressive efforts the social media giant is taking to contain an outbreak of viral falsehoods. (Facebook via AP)

Facebook will soon let you know if you shared or interacted with dangerous coronavirus misinformation on the site, the latest in a string of aggressive efforts the social media giant is taking to contain an outbreak of viral falsehoods.

The new notice will be sent to users who have clicked on, reacted to, or commented on posts featuring harmful or false claims about COVID-19 after they have been removed by moderators. The alert, which will start appearing on Facebook in the coming weeks, will direct users to a site where the World Health Organization lists and debunks virus myths and rumors.

Facebook, Google and Twitter are introducing stricter rules, altered algorithms and thousands of fact checks to stop the spread of bad misinformation online about the virus.

Challenges remain. Tech platforms have sent home human moderators who police the platforms, forcing them to rely on automated systems to take down harmful content. They are also up against people's mistrust of authoritative sources for information, such as the WHO.

"Through this crisis, one of my top priorities is making sure that you see accurate and authoritative information across all of our apps," Facebook CEO Mark Zuckerberg wrote on his Facebook page Thursday.

The company disclosed Thursday that it put more than 40 million warning labels in March over videos, posts or articles about the coronavirus that fact-checking organizations have determined are false or misleading. The number includes duplicate claims—the labels were based on 4,000 fact checks.

Facebook says those warning labels have stopped 95% of users from clicking on the false information.

In this July 30, 2019, file photo, the social media application, Facebook is displayed on Apple's App Store in Chicago, Ill. The social media company said Thursday, April 16, 2020, it is now going to let users know if they liked, reacted or commented on posts with harmful misinformation about the coronavirus that moderators later removed. Facebook says it will also direct people who engaged with those posts to information about virus myths debunked by the World Health Organization. (AP Photo/Amr Alfiky, File)

"It's a big indicator that people are trusting the fact checkers," said Baybars Orsek, the director of the International Fact-Checking Network. "The label has an impact on people's information consumption."

But Orsek cautioned that the data Facebook provided should be

reviewed by outside editors or experts, and called on the historically secretive company to release regular updates about the impact of its fact-checking initiative.

Orsek's organization is a nonprofit that certifies news organizations as fact checkers, a requirement to produce fact-checking articles for Facebook. Facebook has recruited dozens of news organizations around the globe to fact check bad information on its site. The Associated Press is part of that program.

Facebook will also begin promoting the articles that debunk COVID-19 misinformation, of which there are thousands, on a new information center called "Get The Facts." Putting trustworthy information in front of people can be just as useful, if not more, than simply debunking falsehoods.

Still, conspiracy theories, claims about unverified treatments, and misinformation about coronavirus vaccines continue to pop up on the site daily— sometimes circumventing the safeguards Facebook has implemented.

The new notification feature also only applies to posts on users' main news feed—not in groups, where misinformation often spreads unchecked, and not on WhatsApp or Instagram, though Facebook has put some other protections in place on those platforms.

That means a lot of users won't get the new alert from Facebook, said Stephanie Edgerly, an associate professor at Northwestern University who researches audience engagement. She said many users might simply see a false claim in their Facebook feed but not share, like or comment on it.
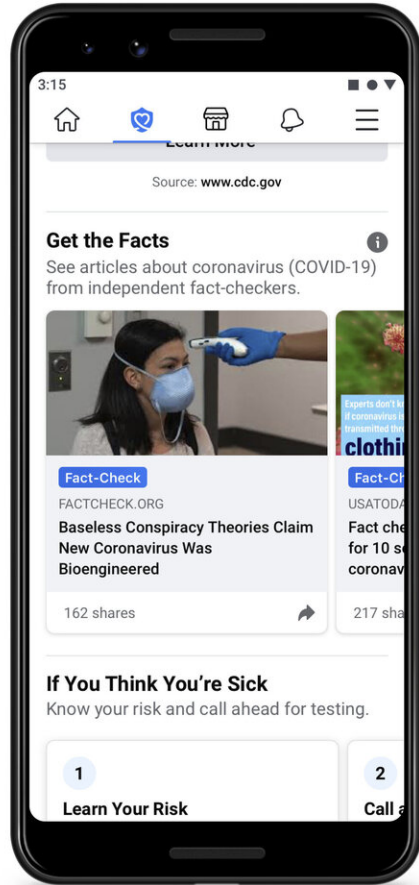
This undated image provided by Facebook shows the Facebook app on a cellular telephone. Facebook will soon let you know if you shared or interacted with dangerous coronavirus misinformation on the site, the latest in a string of aggressive efforts the social media giant is taking to contain an outbreak of viral falsehoods. (Facebook via AP)

"A lot of what we know about how people scroll through their news feed not clicking on things, they still reading posts or headlines, without clicking on the link," Edgerly said.

Facebook users, for example, viewed a false claim that the virus is destroyed by chlorine dioxide nearly 200,000 times, estimates a new

study out today from Avaaz, a left-leaning advocacy group that tracks and researches online misinformation.

The group found more than 100 pieces of misinformation about the coronavirus on Facebook, viewed millions of times even after the claims had been marked as false or misleading by fact checkers. Other false claims were not labeled as misinformation, despite being declared by fact-checkers as false.

"Coronavirus misinformation content mutates and spreads faster than Facebook's current system can track it," Avaaz said in its report.

This is especially problematic for Italian and Spanish misinformation, the report said, because Facebook has been slower to issue warning labels on posts that aren't in English. Avaaz also noted that it can take as long as 22 days for Facebook to label misinformation as such—giving it plenty of time to spread.

False claims about coronavirus treatments have had deadly consequences.

Last month, Iranian media reported more than 300 people had died and 1,000 were sickened in the country after ingesting methanol, a toxic alcohol rumored to be a remedy through private social media messages.

provided for information purposes only.