

Twitter Adds Labels, Warnings to Covid-19 Tweets With Disputed, Misleading Information

|
18 hours ago

Twitter began adding labels and warning messages Monday to tweets containing disputed or misleading information related to the coronavirus pandemic.

The social network said Monday's move marked the next step, following its update in March requiring the removal of tweets that increase the chance of someone contracting or transmitting the virus, including: denial of guidance provided by experts; encouraging people to use fake or ineffective treatments, preventions and diagnostic techniques; and misleading content purporting to be from experts or authorities

Head of site integrity Yoel Roth and director of public policy strategy Nick Pickles said in a blog post that the labels and warning messages will be used to provide additional explanations or clarifications in cases where the risks of harm associated with tweets are less severe, but the content is still misleading or confusing.

The labels and warning messages are similar to those Twitter rolled out in February to address synthetic and manipulated media.

Twitter said the labels—which will also be applied to tweets that were sent before Monday—will link to a Twitter-curated page or an external trusted source to provide users with additional information on the claims being made within the

tweet.

And in cases where the potential for harm is higher, warning labels will be affixed to tweets, alerting people that the information in the tweet contradicts guidance from public health experts.

Roth and Pickles said the social network will take action based on the following three broad categories:

- **Misleading information**—statements or assertions that have been confirmed to be false or misleading by subject-matter experts, such as public health authorities.
- **Disputed claims**—statements or assertions in which the accuracy, truthfulness or credibility of the claim is contested or unknown.

- **Unverified claims**—information (which could be true or false) that is unconfirmed at the time it is shared.

Twitter said it is identifying these tweets via its internal systems to proactively monitor Covid-19-related content, detect high-visibility content and prevent amplification of tweets with labels or warnings, adding that it will continue to rely on its trusted partners to identify content that has the potential to cause offline harm, and it will prioritize the review and labeling of such content.

Roth and Pickles wrote, “We’ll learn a lot as we use these new labels, and are open to adjusting as we explore labeling different types of misleading information. This process is ongoing, and we’ll work to make sure these and other labels and warnings show up across Twitter. Embedded tweets and tweets viewed by people not logged into Twitter may still appear without a label.”

<https://www.adweek.com/digital/twitter-adds-labels-warnings-to-covid-19-tweets-with-disputed-misleading-information/>