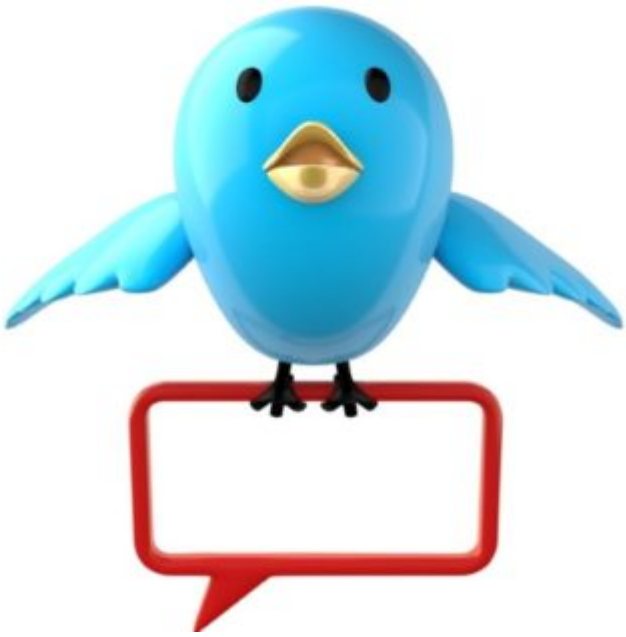
Twitter Releases Its Draft of Potential Rules Covering Deepfakes



|
5 hours ago
Twitter said last month that it would release a draft of its potential rules on deepfakes, and it did so Monday.

The social network said in a series of tweets in October that its new policy will address content that has been "significantly altered or created in a way that changes the original meaning/purpose, or makes it seem like certain events took place that didn't actually happen."

Vice president of trust and safety Del Harvey said in a blog post Monday that Twitter defines synthetic and manipulated media as any photo, audio or video that was significantly altered or fabricated in a way that intends to mislead people or changes its original meaning.

She also shared the following draft of an addition to the Twitter rules for public comment:

Twitter may:

- Place a notice next to tweets that share synthetic or manipulated media;
- Warn people before they share or like tweets with synthetic or manipulated media; or
- Add a link—for example, to a news article or Twitter Moment—so that people can read more about why various sources believe the media is synthetic or manipulated.

In addition, if a tweet including synthetic or manipulated media is misleading and could threaten someone's physical safety or lead to other serious harm, we may remove it.

Harvey said users can share their feedback via this survey—which is available in Arabic, English, Hindi, Japanese, Portuguese and Spanish—or via tweet with the hashtag #TwitterPolicyFeedback.

The feedback period ends Wednesday, Nov. 27, at 11:59 p.m. GMT (6:59 p.m. ET, 3:59 p.m. PT), and Harvey wrote, "At that point, we'll review the input we've received, make adjustments and begin the process of incorporating the policy into the Twitter rules, as well as train our enforcement teams on how to handle this content. We will make another announcement at least 30 days before the policy goes into effect."

She added that people interested in partnering with Twitter to develop solutions that detect synthetic and manipulated media can fill out this form.

https://www.adweek.com/digital/twitter-releases-its-draft-of-potential-rules-covering-deepfakes/