Facebook says it's getting better at removing hate speech

November 15 2018, by Barbara Ortutay

In this Jun 7, 2013, file photo, the Facebook "like" symbol is illuminated on a sign outside the company's headquarters in Menlo Park, Calif. Facebook says it is making progress on deleting hate speech, graphic violence and other violations of its rules, including detecting them before they are seen by users. The company released its second report Thursday, Nov. 15, 2018, detailing how it enforces community standards banning hate, nudity and other content. (AP Photo/Marcio Jose Sanchez, File)

Facebook said it's making progress on detecting hate speech, graphic

violence and other violations of its rules, even before users see and report them.

Facebook said that during the April-to-September period, it doubled the amount of hate speech it detected proactively, compared with the previous six months.

The findings were spelled out Thursday in Facebook's second report on enforcing community standards. The reports come as Facebook grapples with challenge after challenge, ranging from fake news to Facebook's role in elections interference, hate speech and incitement to violence in the U.S., Myanmar, India and elsewhere.

The company also said it disabled more than 1.5 billion fake accounts in the latest six-month period, compared with 1.3 billion during the previous six months. Facebook said most of the fake accounts it found were financially motivated, rather than aimed at misinformation. The company has nearly 2.3 billion users.

Facebook's report comes a day after The New York Times published an extensive report on how Facebook deals with crisis after crisis over the past two years. The Times described Facebook's strategy as "delay, deny and deflect."

Facebook said Thursday it has cut ties with a Washington public relations firm, Definers, which the Times said Facebook hired to discredit opponents. Facebook CEO Mark Zuckerberg said during a call with reporters that he learned about the company's relationship with Definers only when he read the Times report.

On community guidelines, Facebook also released metrics on issues such as child nudity and sexual exploitation, terrorist propaganda, bullying and spam. While it is disclosing how many violations it is catching, the

company said it can't always reliably measure how prevalent these things are on Facebook overall. For instance, while Facebook took action on 2 million instances of bullying in the July-September period, this does not mean there were only 2 million instances of bullying during this time.

In addition, Facebook plans to set up an independent body by next year for people to appeal decisions to remove—or leave up—posts that may violate its rules. Appeals are currently handled internally.

Facebook employs thousands of people to review posts, photos, comments and videos for violations. Some things are also detected without humans, using artificial intelligence. Zuckerberg said creating an independent appeals body will prevent the concentration of "too-much decision-making" within Facebook.

Facebook has faced accusations of bias against conservatives—something it denies—as well as criticism that it does not go far enough in removing hateful content.