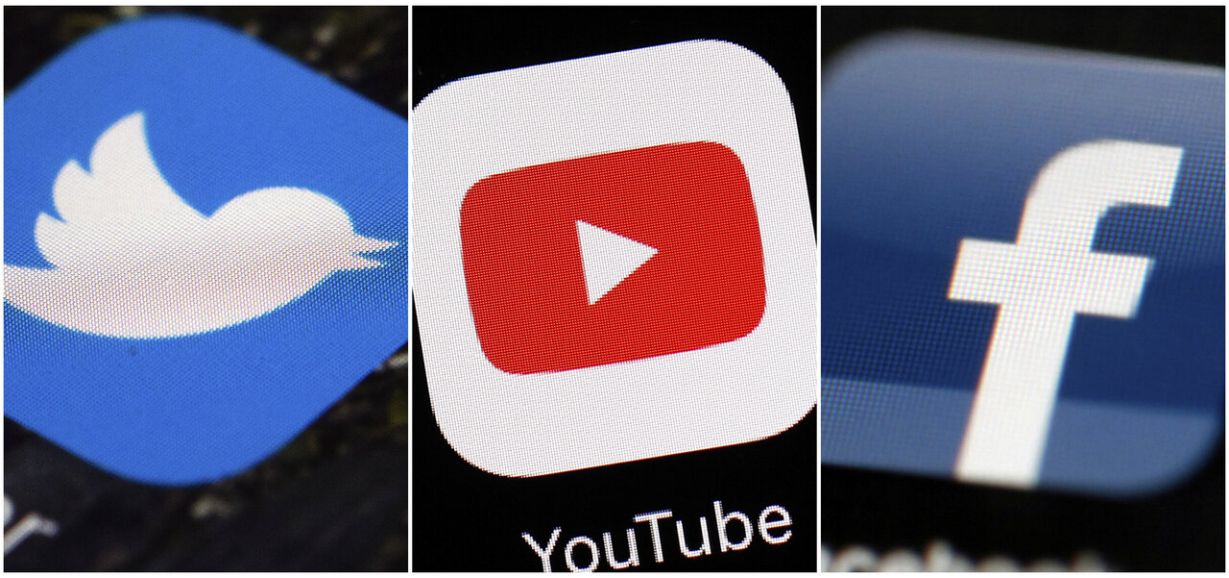
NATO researchers: Social media failing to stop manipulation

December 6 2019, by Kelvin Chan



This combination of images shows logos for companies from left, Twitter, YouTube and Facebook. Social media companies are failing to stop manipulated activity, according to a report Friday, Dec. 6, 2019 by NATO-affiliated researchers who said they were easily able to buy tens of thousands of likes, comments and views on Facebook, Twitter, YouTube and Instagram. Most of the phony accounts and the activity they engaged in remained online weeks later, even after researchers at the NATO Strategic Command Centre of Excellence flagged it up as fake. (AP Photos/File)

Social media companies are failing to stop manipulated activity, according to a report Friday by NATO-affiliated researchers who said

they were easily able to buy tens of thousands of likes, comments and views on Facebook, Twitter, YouTube and Instagram.

Most of the phony accounts and the activity they engaged in remained online weeks later, even after researchers at the NATO Strategic Command Centre of Excellence flagged them up as fake.

The center, an independent group based in Latvia that advises the military alliance, said the findings contrast with statements from tech companies that say they've been working harder on stamping out manipulation.

"Overall social media companies are experiencing significant challenges in countering the malicious use of their platforms," the report said.

Online manipulation emerged as a major issue for tech companies after the 2016 U.S. election, when Russian influence efforts came to light. The researchers found that most fake social media activity is bought for commercial, not political, reasons. It can include Instagram influencers trying to pump up their profiles to make more money from their brand contracts.

Fake accounts are still used for political means, though it's a minor slice of the industry and aimed at "non-western" pages, the researchers said, noting they were used to buy engagement on hundreds of political pages and dozens of government pages.

To carry out the study, the researchers turned to the "manipulation service provider" industry, which is expanding to feed the growing demand for phony clicks and likes. They used 16 companies, most based in Russia, to buy fake online engagement for 105 posts on Facebook, Twitter, YouTube and Instagram. They spent just 300 euros ($330) to purchase 3,530 comments, 25,750 likes, 20,000 views and 5,100

followers.

To avoid influencing real conversations, they only bought clicks for posts that were at least six months old and carried neutral and non-political messages, such as "Hello!" and "Thank you!" on New Year's greetings from European Union commissioners.

Four weeks later, 80% of the fake activity remained online, the researchers found, as they sought to gauge whether the sites were independently detecting misuse. They then reported 100 of the accounts as fake, but found about 95 remained active three weeks later.

Some companies were better than others, the report said.

YouTube was the easiest site on which to create fake accounts but the best at countering artificial likes and video views. Manipulating Instagram is easy and cheap because the site is was largely unable to detect and stop it, while Twitter was best at detecting and removing manipulation.

Facebook was best at stopping fake accounts, but any that got through were more successful because they faced little further scrutiny, and their comments and views weren't removed. Facebook says it disabled 2.2 billion fake accounts in the first quarter of this year.

"Fake engagement tactics remain a challenge facing the entire industry," Facebook, which also owns Instagram, said in a statement. "We're making massive investments to find and remove fake accounts and engagement every day."

YouTube said it takes any abuse of its systems seriously and has invested in technology to prevent the artificial inflation of video view counts.

"While no anti-spam system will ever be perfect, our teams work very hard to manage spam views to less than one percent of all views," it said in a statement.

Twitter said it has "invested significant technical resources to this issue and are committed to improvement."