July 31, 2019

Honorable Adam Schiff
United States House of Representatives
2269 Rayburn House Office Building
Washington, DC 20515

Dear Representative Schiff,

Thank you for your letter of July 15, 2019, regarding a doctored video of Speaker Nancy Pelosi that was shared on Facebook in May of this year. We appreciate the opportunity to address your questions.

Facebook is a community of more than two billion people, spanning countries, cultures, and languages across the globe. Every day, the members of our community express themselves on our platform in diverse ways, having conversations and posting everything from text to links to photos and videos. We encourage and are proud of the wide array of expression on Facebook, but also recognize that we have an important role to play in keeping abuse off our platform.

To that end, Facebook has Community Standards that govern what may and may not be on Facebook, and Advertising Policies that govern what may be advertised. For example, among the content we disallow is hate speech, content posted by fake accounts, and harassment. When we become aware of content that violates these policies—either through proactive technical measures or through user reports—we remove it.

There is some content that lacks quality and integrity despite not violating our Community Standards. For example, in the case of misinformation, rather than remove it, we work to reduce its distribution on the platform and to inform users with additional context so they can decide whether to read, trust, or share it. This is how we handle all misinformation, including manipulated photos and video, on the platform.

People share millions of photos and videos on Facebook every day. We know that this kind of sharing is particularly compelling because it is visual. That said, it also creates an opportunity for manipulation by bad actors. In 2018, we expanded our photo and video fact-checking capabilities to all of our third-party fact-checking partners to help us identify and take action against more types of misinformation, including manipulated photos and videos, more quickly. More recently, we announced partnerships with academics to invest in new techniques to detect manipulated media across images, video, and audio.

Leading up to the 2020 U.S. presidential election cycle, we know that combating misinformation, including deepfakes, is one of the most important things we can do. We continue to look at how we can improve our approach and the systems we've built, including through engagement with academics, technical experts and policymakers.

With that context in mind, please find answers to your specific questions below.

1.) How many views did the manually altered video of Speaker Pelosi receive on Facebook before it was marked as "false" by independent fact checkers? How long did it take to initiate and then complete that independent review? How many views did the video receive after being marked false?

facebook

The video was posted to Facebook on Wednesday, May 22, at 10:29 am PT. The video was rated false by our third-party fact-checkers on the afternoon of Thursday, May 23, after which the video's distribution was dramatically reduced.

Lead Stories, one of the independent third-party fact-checkers with whom Facebook works, published a public analysis of the impact of their fact-check of the video and concluded "[a]t the time Lead Stories flagged it as false, the video had 46,519 shares, 8,692 likes, and 2,268,188 views. Soon after we published, the peak rate of 2,124 shares per hour measured just two hours earlier was throttled to a trickle of just 43 shares an hour... Four days after Lead Stories first flagged the video as false—triggering all of the actions designed by Facebook to limit the spread of fake content--it was shared less than 2,000 times, a very slow average rate of 20 times per hour. It gained less than 3,000 additional likes, an average of just 43 per hour. It was viewed only 570,000 times more in those subsequent four days, a not-so-hot rate of 6,000 per hour."

2.) In a recent interview, you stated that deepfakes may be "a completely different category of thing from normal false statements overall." Does Facebook have a written policy on deepfake content on Facebook, Instagram, and WhatsApp? If so, will you provide it in response to this letter? If not, are you developing such a policy and when will it be finalized?

Deepfakes and other forms of manipulated media are governed by the same policies and procedures outlined above for misinformation: we remove content that violates our Community Standards and reduce the distribution of false or misleading content on our platforms.

Deepfakes take a number of different forms. Some of this content runs afoul of our existing content policies. For example, a synthesized video of a celebrity in which the celebrity is nude would violate our nudity policies. Deepfakes may also be spread by inauthentic accounts, which would violate our policies; in such cases, the content posted by such accounts would also be removed. Deepfakes are also eligible for fact-checking and News Feed reduction in Facebook and Instagram Explore. Many of our third-party fact-checking partners have expertise evaluating photos and videos and are trained in visual verification techniques, such as reverse image searching and analyzing the image metadata that indicates when and where the photo or video was taken. Fact-checkers are able to assess the truth or falsity of a photo or video by combining these skills with other journalistic practices, including by using research from technical experts, academics or government agencies.

As you note, Mark Zuckerberg recently posed the question of whether deepfakes are different from other false content, and he shared that we are currently considering this question in our policy development process. This process includes defining deepfakes with precision so that we can apply our rules consistently, and talking to internal teams and external experts. We have recently engaged with more than 50 global experts with technical, policy, media, legal, and academic backgrounds to inform our policy development process. For example, we recently participated in a synthetic and manipulated media roundtable hosted by the Carnegie Endowment for International Peace with some of the world's topmost experts, contributed to a manipulated media workshop led by the BBC and WITNESS (a leading civil society organization focused on the use of video and technology to protect human rights), and have an ongoing collaboration with the Partnership on AI.

facebook

We continue to discuss these issues with internal and external stakeholders. Based on these discussions, we are considering a number of options to better address manipulated media and deepfakes, including developments related to policies, collaboration on detection, and more. We commit to publicly announce any significant changes resulting from these consultations.

3.) Facebook's Terms of Service explicitly prohibit users from sharing "anything that is unlawful, misleading, discriminatory or fraudulent." Are fake images or videos that realistically portray individuals saying or doing something they never did considered to be misleading? Are they allowed on Facebook's platforms? Under what circumstances, if any, would Facebook remove such content and block its upload to your platforms?

Facebook's Terms of Service identify broadly the types of content that Facebook can enforce against but do not specify how enforcement will apply. Content removal enforcement is governed by the Community Standards and Advertising Policies described above. For example, the term "misleading" in the Terms is intended to capture things like SPAM or other commercial content that misleads consumers, misinformation related to voter suppression, or coordinated inauthentic behavior. In those contexts, which are laid out in the Community Standards, we remove misleading content.

As described above, we would remove content such as you describe if it ran afoul of any of our Community Standards. Additionally, we recently updated our policies to address misinformation for content monetization; any content marked as false or misleading by third-party fact-checkers is now automatically made ineligible for monetization.
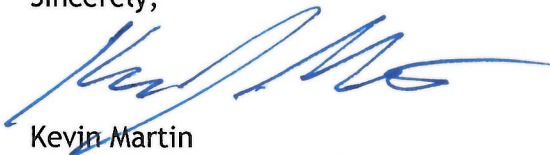
4.) Is Facebook conducting research into techniques for automatically detecting deepfakes and other forms of machine-manipulated media on its platforms? To the extent machine-manipulated media is detected upon upload to a Facebook platform, will Facebook take specific steps to dampen the virality of such content, take it down completely, or require a human review for politically relevant content?

We have invested heavily in hiring to promote the integrity, safety, and security of our platform and are fortunate to have among our ranks some of the world's foremost experts in AI and machine learning. These experts are leading company efforts to enhance detection of manipulated media. In addition, we are investing in and committed to academic partnerships with the University of Maryland, Cornell University, and The University of California, Berkeley, (and exploring additional collaborations) to research new techniques to detect manipulated media across images, video, and audio. Dealing with the rise of manipulated media will require deep research and collaboration between industry and academia—we need everyone working together to tackle this challenge. These partnerships are only one piece of our efforts to partner with academics and our colleagues across industry—in the months to come, we will partner more so we can all move as quickly as possible to innovate in the face of this concern. This work will be critical for our broader efforts against manipulated media, including deepfakes. These are complex issues in an adversarial space where techniques and tactics are rapidly evolving. We know that it is only by remaining vigilant and working with experts and other companies, governments, and civil society around the world that we will be able to stay ahead of these new means of generating false content.

As we evaluate any updates to our policies and practices in the coming months, these investments in better technical detection will be crucial; we are committed to continually and iteratively continuing to invest, engage, collaborate, and improve detection as this technology advances.

Thank you again for the opportunity to address these questions. We appreciate your leadership on this issue and applaud your work to include a deepfake prize competition in the Intelligence Authorization Act. It is going to take government, academics, and the private sector working together to develop the tools necessary to combat manipulated media. We look forward to working with you and having an open dialogue as we continue to evaluate our policies and how best to address these issues.

Sincerely,

Kevin Martin
Vice President, U.S. Public Policy