July 24, 2020

Chairman Adam Schiff
Permanent Select Committee on Intelligence
U.S. House of Representatives
Capitol Visitor Center HVC-304
Washington, D.C. 20515

Dear Chairman Schiff and Members of the Committee:

Thank you for your questions for the record from the June 18, 2020 virtual hearing entitled Emerging Trends in Online Foreign Influence Operations: Social Media, COVID-19, and Election Security. Per your request, attached are the answers for the record to your questions.

Sincerely,

Facebook, Inc.

575 7TH STREET NW
STE 700
WASHINGTON, D.C.
20004

Questions from Chairman Schiff

For all witnesses

1. **In the course of removing assessed networks engaged in CIB or foreign influence operations, does your company having standing policy or guidance with respect to proactively informing users who engaged with those removed accounts or the content? Why or why not?**

We have worked to notify people about foreign influence operations on a variety of occasions and will continue to do so as appropriate. Over the past three years, Facebook has publicly shared information about the coordinated inauthentic behavior we detect and remove from our platforms. In February, we consolidated our public announcements to a consistent monthly report, to ensure the public can find up-to-date analysis of the deceptive behavior we are seeing and actioning. As part of our regular reports, we're sharing information about all networks we take down over the course of a month to make it easier for people to see the progress we're making in one place. For more information, please visit https://about.fb.com/news/tag/coordinated-inauthentic-behavior/.

2. **Can you please describe your company's relationships or engagements with the national political parties, state parties, and individual campaigns, generally, and in the event you discovered a covert foreign influence operation targeting a specific candidate or political party?**

   a) **Are these interactions regular, or would they depend on identification of a specific threat?**

   b) **If an individual candidate suspects they are being subjected to malign online activity, do they know who and how to contact at your company?**

If we find instances of coordinated inauthentic behavior conducted on behalf of a foreign actor, regardless of whether or not such behavior targets a candidate or political party, we apply the broadest enforcement measures, including the removal of every on-platform property connected to the operation itself and the people and organizations behind it. We also report publicly about such takedowns in a monthly report, available at https://about.fb.com/news/tag/coordinated-inauthentic-behavior/.

Regarding our efforts to protect campaigns and candidates, last year we launched Facebook Protect to further secure the accounts of candidates, elected officials, federal and state departments and agencies, and party committees in the US, as well as their staff. As we've seen in past elections, they may be particularly vulnerable to targeting by hackers and foreign adversaries. However, because campaigns are generally run for a short period of time, we do not always know who these campaign-affiliated people are, making it harder to help protect them.

Facebook Protect allows Page admins to enroll their organization's Facebook and Instagram accounts and invite members of their organization to participate in the program as well. Participants will be required to turn on two-factor authentication, and their accounts will be monitored for hacking, such as login attempts from unusual locations or unverified devices. And

if we discover an attack against one account, we can review and protect other accounts affiliated with that same organization that are enrolled in our program. You can find more information about Facebook Protect at https://www.facebook.com/gpa/facebook-protect.

3. We've seen China in particular engage in overt use of its official diplomatic accounts and state-controlled media to shape the information space online and promote misleading or false narratives that advance its state strategic interests in an identifiably coordinated manner. Beyond mere labeling of state-controlled media or identification of official foreign or diplomatic account as such:

a) Can you please describe your company's approach to fact-checking or adding context to misleading or outright disinformation posted by these overt, foreign- linked accounts in a coordinated manner, which might allow users to readily understand the broader context or be directed to authoritative, credible sources about the claims?

If we find instances of coordinated inauthentic behavior conducted on behalf of a government entity or by a foreign actor, in which the use of fake accounts is central to the operation, we apply the broadest enforcement measures, including the removal of every on-platform property connected to the operation itself and the people and organizations behind it.

When it comes to misinformation from authentic accounts, we partner with over seventy fact-checking organizations around the world that fact-check content in more than fifty languages to combat misinformation and reduce the spread of false news. If content is deemed by a fact-checker to be false or partly false, its distribution will be reduced and it will appear lower in News Feed. We also implement an overlaid warning screen on top of content marked as false. People who try to share the content will be notified of the fact-checker's reporting and rating and they will also be notified if content they have shared in the past has since been rated false by a fact-checker. We also take action against Pages and domains that repeatedly share or publish content that is rated "false." Such Pages and domains will see their distribution reduced as the number of offenses increases. Finally, Pages and domains that repeatedly publish or share false news will also lose their ability to register as a news Page on Facebook, and if a registered news Page repeatedly shares false news, its news Page registration will be revoked.

b) If a Facebook post, Tweet, or YouTube video created by a state-controlled media outlet promotes misleading or provably false narratives in apparent coordinated manner reasonably assessed to be in the service of that state's interests, what steps might your respective platforms consider in terms of labeling, fact-checking, or providing context to users about such material?

As discussed in the answer above, if we find instances of coordinated inauthentic behavior conducted on behalf of a government entity or by a foreign actor in which the use of fake accounts is central to the operation, we apply the broadest enforcement measures, including the removal of every on-platform property connected to the operation itself and the people and organizations behind it. We regularly share our findings about the networks we find and remove for coordinated inauthentic behavior.

Facebook labels media outlets that we believe are wholly or partially under the editorial control of their government. We provide greater transparency into these publishers because they combine the influence of a media organization with the strategic backing of a state, and we believe people should know if the news they read is coming from a publication that may be under the influence of a government. And to ensure we're equally transparent when it comes to paid content from these publishers, we will begin labeling ads from these publishers later this year. The labels will appear globally in the Ad Library Page view, on Pages, and in the Page Transparency section. In the US, the label appears on posts in News Feed.

Last month, we began blocking ads from entities that Facebook has designated as state media outlets in the US out of an abundance of caution to provide an extra layer of protection against various types of foreign influence in the public debate ahead of the November 2020 election.

Regarding Facebook's efforts to fact-check such content, please see the answer to your Question 3(a), above.

4. Graphika's June 16, 2010 report about the so-called "Secondary Infektion" group assessed it as having links to Russia and attempted to use false stories and outright forged materials to advance narratives favorable to Moscow.

a) Does your company have a policy governing the removal of "genuine," provably hacked or stolen materials found on your platform, similar to the episode involving the hacked-and-dumped emails of Clinton Campaign Chair John Podesta in 2016? If so, please provide it in writing.

Yes. We prohibit any content that is claimed or confirmed to have come from a hacked source, regardless of whether the affected person is a public figure or a private figure. In rare situations and on a case-by-case basis, we may choose to allow content that is newsworthy, significant, or important to the public interest even if it otherwise violates our policies. We do this only after weighing the public interest value of the content against the risk of real-world harm.

b) Does this policy include or account for the posting of suspected or proven forgeries that were presented as genuine and was linked to a foreign influence operation? Or would your company otherwise prevent the sharing or re-posting of such forged content?

We investigate and enforce against any type of inauthentic behavior. If we find instances of coordinated inauthentic behavior—whether conducted by foreign or domestic actors—we remove both the accounts and the content connected to the operation itself and the people and organizations behind it.

Content shared authentically (not coming from a hacked source), but that is still a forgery or misrepresentation, is eligible for rating by our fact-checking partners. When they rate content as false or partly false, we reduce its distribution, implement an overlaid warning screen on top of the content, notify people who try to share the content or have shared the content in the past, and reject its inclusion in ads.

c) Do these or other policies cover content that might otherwise be illicitly obtained, e.g. a phone conversation that was recorded by a third party without the knowledge or consent of the calling or the called party, and then posted to Facebook, Twitter, or YouTube?

Generally speaking, Facebook prohibits people from facilitating organizing, promoting, or admitting to certain criminal or harmful activities. This includes statements of intent, calls to action, or advocating for hacking or theft. For more information, see https://www.facebook.com/communitystandards/coordinating_harm_publicizing_crime.

5. What changes has your company made to algorithms deployed on its internet platforms since 2017, especially with respect to limiting the reach or potential virality of extremist content and conspiracy theories?

a) How do you measure your success?

b) Would you make public metrics so that we in Congress can judge these issues in a non-anecdotal fashion?

Content that violates our Community Standards, including terrorist activity, organized hate, bullying and harassment, calls for violence, and more, is removed from Facebook.

The News Feed algorithm looks at thousands of signals to show the most relevant and meaningful content to each individual. A person's News Feed is not static, but rather is personalized based on signals such as the user's Facebook activity—for example, likes, comments, the Pages they follow, and who their friends are. Users who do not wish to consume the algorithmically ranked version of News Feed also have the option to view content chronologically from those they follow in the 'Most Recent' Feed view. For more information, please see https://www.facebook.com/help/218728138156311.

We frequently make changes to our algorithms in an effort to improve people's experience on Facebook. For example, in 2018, we responded to feedback from our community that public content—posts from businesses, brands, and media—was crowding out the personal moments that lead us to connect more with each other. As a result, we moved from focusing only on helping users find relevant content to helping them have more meaningful social interactions. This meant that users began seeing more content from their friends, family, and Groups. We also reduce the distribution of some problematic types of content, including content that users may find spammy or low-quality, such as clickbait headlines and links to low-quality webpages like ad farms.

Separately, we also work to reduce the spread of viral misinformation, which can include conspiracy theories, on our platform. We work with independent, third-party fact-checkers to help reduce the spread of false news and other types of viral misinformation. If content is deemed by a fact-checker to be false or partly false, its distribution will be reduced, and it will appear lower in News Feed. We also implement an overlaid warning screen on top of content marked as false. People who try to share the content will be notified of the fact-checker's reporting and rating and they will also be notified if content they have shared in the past has since been rated false by a fact-checker. We also take action against Pages and domains that

repeatedly share or publish content that is rated "False." Such Pages and domains will see their distribution reduced as the number of offenses increases, including their eligibility for recommendations and ability to advertise and monetize. Finally, Pages and domains that repeatedly publish or share false news will also lose their ability to register as a news Page on Facebook, and if a registered news Page repeatedly shares false news, its news Page registration will be revoked.

To track our progress and demonstrate our continued commitment to making Facebook safe and inclusive, we regularly release our Community Standards Enforcement Report (available at https://transparency.facebook.com/community-standards-enforcement). This report shares metrics on how Facebook is performing in removing content that goes against our Community Standards. We also release a "prevalence" metric that estimates how much violating content has been posted on the platform. We share data on our process for appealing and restoring content to correct mistakes in our enforcement decisions.

For Facebook

1. What overall outcomes and metrics do Facebook's content prioritization algorithms (e.g. News Feed) optimize for?

a) How does user engagement rank among these factors, and how is it evaluated?

b) Are measures of accuracy, veracity, reliability, authenticity, divisiveness, or original posts from family and friends taken into account, and how are such factors weighted?

c) How has Facebook's approach changed since the tendency of its algorithms to amplify extremist or polarizing content became apparent in the wake of 2016?

d) Please provide specific details about the "number of changes" Mr. Gleicher cited in his testimony.

People see posts from their friends, Pages they've chosen to follow, and Groups they've joined, among others, in their News Feed. On a given day, the number of eligible posts in a user's Feed inventory can number in the thousands, so we use an algorithm to personalize how this content is organized. The goal of the News Feed algorithm is to predict what pieces of content are most relevant to the individual user, and rank (i.e., order) those pieces of content accordingly every time a user opens Facebook, to try and bring those posts that are the most relevant to a person closer to the top of their News Feed. This ranking process has four main elements: the available inventory—all of the available content from the people, Pages, and Groups a person has chosen to connect with; the signals, or data points, that can inform ranking decisions, e.g., who posted a particular piece of content; the predictions we make, including how likely we think a person is to comment on a story, share with a friend, etc.; and a relevancy score for each story.

We frequently make changes to the algorithms that drive News Feed ranking in an effort to improve people's experience on Facebook. For example, in 2018, we responded to feedback from our community that public content—posts from businesses, brands, and media—was crowding out the personal moments that lead us to connect more with each other. As a result, we moved from focusing only on helping users find relevant content to helping them have more meaningful social interactions. This meant that users began seeing more content from their friends, family, and Groups. We also reduce the distribution of some problematic types of content, including content that users may find spammy or low-quality, such as clickbait headlines and links to low-quality webpages like ad farms.

To help people on Facebook better understand what they see from friends, Pages, and Groups in News Feed, including how and why that content is ranked in particular ways, we publish a series of blog posts called Newsroom posts, which highlight major updates to News Feed and explain the thinking behind them. Also, in 2019, we launched a feature called "Why am I seeing this post?" (see https://about.fb.com/news/2019/03/why-am-i-seeing-this/). This feature directly responded to user feedback asking for more transparency around why certain content appears in News Feed and easier access to News Feed controls. Through their News Feed Preferences, users can choose to see posts from certain friends and Pages higher up in their News Feed. Controls also include Snooze, which keeps the content from a selected person, Page, or Group out of a user's News Feed for a limited time.

Users who do not wish to consume ranked News Feed also have access to a control to view content purely chronologically from those they follow in the 'Most Recent' Feed view (see https://www.facebook.com/help/218728138156311). Additionally, we promoted a series of educational initiatives and campaigns to help people learn about the technology that underlies our various products and features, which includes AI and machine learning, through our series called "Inside Feed" (see https://about.fb.com/news/category/inside-feed/).

Facebook is a platform that reflects the conversations already taking place in society. We are keenly aware of the concern that our platform is contributing to polarization, and we have been working to understand the role that we play in discourse and information diversity. The data on what causes polarization and "filter bubbles" is mixed. Some independent research has shown that social media platforms provide more information diversity than traditional media, and our own research indicates that most people on Facebook have at least some friends who claim an opposing political ideology—probably because Facebook helps people maintain ties with people who are more distantly connected to them than their core community—and that the content in News Feed reflects that added diversity. We want Facebook to be a place where people can discover more news, information, and perspectives, and we are working to build products that help. And, because we want Facebook to be a place where people can express themselves, we must also preserve our community's sense of safety, privacy, dignity, and authenticity via our Community Standards, which define what is and isn't allowed on Facebook and Instagram. We remove content that violates our Community Standards, such as hate speech, bullying, and harassment.

2. What factors do Facebook's algorithms take into account when ranking or surfacing News Feed items, suggesting videos or groups, or otherwise prioritizing or recommending user-generated content?

a) Which are given the most weight on average, and how do factors related to engagement (e.g. number of likes or comments, etc.) rank?

b) Are any based on predicted performance, either based on past behavior of the creator, sharer, or viewer, or based on the substance of the content itself?

Please see the response to your Question 1.

3. The June 16, 2020 Graphika report on "Secondary Infektion" begins a with description of what the report describes as a small cluster of accounts linked to Russian actors in May 2019. The report states that Facebook made that attribution based on technical signals. Without revealing information that could compromise your investigative tools:

a) What can you share with us about what the indicators that enabled you to make attribution to Russian-linked actors in this case?

Our investigative team first discovered Secondary Infektion, and then shared information with Graphika and our partners in industry. It has been encouraging to see the broad societal response to these actors since they were first exposed, including by other tech platforms and researchers.

Determining attribution to a specific organization or entity is challenging for a private sector company; it is especially hard without access to the type of information that governments can use to determine attribution. At Facebook, we look at a variety of signals, using a mix of our technology and investigative work. In doing that, we try to link suspicious activity to individuals or entities with primary operational responsibility for the malicious action.

In this case, we removed 21 Facebook accounts, Pages, and Instagram accounts that were involved in coordinated inauthentic behavior as part of a small network emanating from Russia that focused on Austria, the Baltics, Germany, Spain, Ukraine, and the United Kingdom. We saw this network was also active on other internet platforms and engaged in a number of deceptive tactics, including the use of fake accounts to join Groups, impersonate other users, and amplify allegations about a public figure working on behalf of intelligence services. They also posted content about local politics, including topics like immigration, religious issues, and NATO. From what we saw on our platforms, this operation prioritized operational security which led to its inability to gain much following.

This investigation, perhaps more than others, has made it clear that these operations are rarely confined to one platform. That's why we're working closely with other tech companies to deal with the threats we're seeing. As we normally do, we shared information about our findings with industry peers, law enforcement in relevant countries, and researchers. Our takedown in May 2019 kicked off a series of investigations around the world by researchers and other platforms.

b) Have you shared information about those specific technical indicators with US government partners focused on foreign interference? Why or why not?

Yes, we reported this takedown. Information about this takedown is available at https://about.fb.com/news/2019/05/more-cib-from-russia/.

We have a long history of working successfully with law enforcement, including the FBI and DHS, to address foreign and domestic influence operations. In this case, we shared information about our findings with law enforcement.

4. As a hypothetical example, were Facebook to identify with high confidence a network of accounts tied to a foreign influence operation targeting the United States:

a) Do those accounts enjoy the same rights under your terms of service as authentic users?

b) If not, and if you determine the foreign influence network in this scenario is in violation, are you then able to maximally share all associated indicators, metadata, or even content with other companies or the U.S. Government? Why or why not?

Inauthentic behavior, including foreign influence campaigns, has no place on Facebook. If we find any instances of coordinated inauthentic behavior targeting the US conducted on behalf of a foreign actor, we apply the broadest enforcement measures, including the removal of every on-platform property connected to the operation itself and the people and organizations behind it. So far this year, we have taken down twenty-two networks that were engaging in this sort of deceptive behavior, including three networks originating from Russia, two from Iran, and three based here in the United States.

We know that inauthentic behavior is not limited to a specific type of technology or service. The better we can be at working together with industry and outside security researchers, the better we'll do by our community. We continuously look for ways to enhance our collaboration with industry and the security research community while ensuring that we put the right checks in place to protect people's information.

That's why we're working closely with our fellow tech companies to deal with the threats of inauthentic behavior we have all seen. Several takedowns that we conducted and announced were in close collaboration with other tech platforms, security companies, and law enforcement agencies. For instance, in March 2020, we took down a network of accounts engaging in foreign interference in Ghana and Nigeria on behalf of Russia targeting primarily the United States. We shared this information with other tech companies, including Twitter, which also announced the takedown of this activity on its platform. For more information, see https://about.fb.com/news/2020/03/removing-coordinated-inauthentic-behavior-from-russia/.

We are also committed to working with law enforcement, and we deeply respect and support the work law enforcement agencies do to keep us safe. We have a long history of working successfully with law enforcement, including the FBI and DHS, to address foreign and domestic influence operations.

5. Facebook recently began releasing monthly reports about the removals of coordinated inauthentic manipulation of the platform by foreign actors, which is a positive development.

a) Is it Facebook's position that it is already sufficiently, maximally sharing all relevant data, indicators, with third-party research groups or academics, even if it's not sharing such information with the public at large? Why or why not?

We were the first platform to regularly issue updates on our work against influence operations and we share our findings about the networks we remove for engaging in deceptive behavior on our platforms. Over the past three years, we've shared information about the coordinated inauthentic behavior we detect and remove from our platforms. In February, we consolidated our public announcements to a consistent monthly report, to ensure the public can find up-to-date analysis of the deceptive behavior we are seeing and actioning. As part of our regular reports, we're sharing information about all networks we take down over the course of a month to make it easier for people to see the progress we're making in one place. For more information, please visit https://about.fb.com/news/tag/coordinated-inauthentic-behavior/.

In addition, we partner with independent research teams, including at the Atlantic Council's Digital Forensic Research Lab, Stanford's Internet Observatory, Graphika, and others to share our findings so they can provide additional independent analysis of the coordinated inauthentic behavior we identify, take down, and publicly share, including networks' behavior off-platform and across different internet services.

For more information about our work with law enforcement and our industry partners, please see the response to your Question 4 above.

b) Is Facebook considering releasing more detailed information publicly – such complete lists of the fake accounts or pages, or compendiums of posted content – it has removed in the course of a foreign influence investigation? Why or why not?

As a matter of policy, we currently do not publicly release the full set of Pages and accounts involved in these takedowns. We share examples of posts that cover a wide range of topics in various countries targeted by these activities to help inform the public about what we've found, while being cautious about protecting people's privacy and safety. We want to be careful in sharing this information because we don't want to involve innocent people, who may have been swept up in these campaigns unwittingly. We mitigate these risks by sharing with experts and researchers so we can ensure public awareness and independent analysis without putting people's information at risk.

We partner with researchers at the Atlantic Council's Digital Forensic Research Lab, Graphika, Stanford's Internet Observatory, and others. These experts provide additional analysis of the coordinated inauthentic behavior we identify, remove, and publicly share, including networks' behavior off-platform and across different internet services.

c) Will Facebook release a comprehensive index of the 2016-2017 organic content it attributed to the Russian IRA, properly redacted for privacy protection as necessary, to allow researchers and the public to see how this activity was intended to influence the political conversation during the last presidential election? Why or why not?

Facebook has already provided these posts to the House Permanent Select Committee on Intelligence, the Senate Select Intelligence Committee, and the Senate Judiciary Committee. We do not plan to release a public index. We have, however, provided redacted copies of the advertisements identified on Facebook as having been generated by the Internet Research Agency from 2015, 2016, and 2017. This content is available to the public at https://intelligence.house.gov/social-media-content/social-media-advertisements.htm.

d) What limits in law or regulations exist that would otherwise prevent Facebook from sharing more with: 1) outside experts on a confidential basis; 2) with affected users; or 3) the public at large?

As Nathaniel Gleicher testified, information sharing among the industry and the government has improved over the past few years. And we work closely with law enforcement, industry partners, and civil society. That said, the industry would benefit from a clear legal framework regarding data sharing in the context of investigating influence operations.

We continuously look for ways to enhance our collaboration with industry and the security research community while ensuring that we put the right checks in place to protect people's information, because we know that inauthentic behavior is not limited to a specific type of technology or service. The better we can be at working together with industry and outside security researchers, the better we'll do by our community.

6. How does the use or potential implementation of encryption on platforms like WhatsApp inhibit your ability to detect foreign influence activities? How do you plan to mitigate such limitation, if that's even possible? And how do you evaluate the tradeoffs involved?

It's important to note that our work against influence operations or coordinated inauthentic behavior on our platforms is focused on violating behavior rather than content being shared by these networks. Because of that, we look at patterns of activity and coordination across assets, rather than the content they share. That approach allows us to respond to coordinated manipulation campaigns globally.

For WhatsApp, we take a hard line against bulk or mass messaging and we ban over 2 million accounts per month by relying on advanced machine learning systems to detect coordinated abuse of this nature. We also rely on user reports to evaluate the potential motivation behind the abuse to further improve our systems for the future—to prevent automated messaging or activities that might have an economic or political agenda, for example.

Strong encryption provides numerous security benefits to users, including protecting them from messages in the event of a server-side compromise, like we have witnessed with other technology companies over the years.

We also want to be clear that it is not the intent of our product and service changes to diminish or adversely affect our ability to work with law enforcement and national security authorities, including in our efforts to detect coordinated inauthentic behavior. On WhatsApp, which is an encrypted messaging service, we rely on all available unencrypted information, including profile photos and group information, to detect and prevent coordinated inauthentic behavior.

We are working on developing the strongest techniques for safety within the framework of end-to-end encrypted messaging services. As we move to end-to-end encryption across our messaging platforms, these capabilities to detect bad actors will only get stronger as we are able to obtain additional signals from the public portions of our platform.

7. What visibility does Facebook have into private groups when it comes to enforcing against CIB, foreign influence activity, or violations of other Facebook policies that relate to misinformation, harmful conspiracies, incitements of violence, threats of physical harm, or manipulated media?

a) Have private groups factored into any of Facebook's investigations and removals of CIB activity from 2019 or 2020?

b) Have indicators or metadata from private Facebook groups been given to U.S. agencies in the context of potential criminal activity, threats of violence, or foreign influence operations?

When we take down influence operations, we remove all their assets, often including Groups.

Private Groups can be important places for people to come together and share around a range of personal topics, but being in a private Group doesn't mean that a user's actions should go unchecked. We have a responsibility to keep Facebook safe, which is why our Community Standards apply across Facebook, including in private Groups. We have a specialized team that has been working on the Safe Communities Initiative, which seeks to protect people using Facebook Groups from harm. Made up of product managers, engineers, machine learning experts, and content reviewers, this team works to anticipate the potential ways people can cause harm in groups and develop solutions to minimize and prevent it.

In terms of enforcing the Community Standards in private Groups, we focus on detecting violating content proactively, providing tools for Group admins, and providing transparency and control for Group members. When it comes to detection, we use AI and machine learning to proactively detect bad content before anyone reports it, and sometimes before people even see it. As content is flagged by our systems or reported by people, trained reviewers consider context and determine whether the content violates our Community Standards. We then use these examples to train our technology to get better at finding and removing similar content. This process applies to all public and private Groups.

To help admins run meaningful Groups, we built Group Quality, which gives admins an overview of content Facebook has removed and flagged to them for most Community Standards violations. We also added a section about false news found in Groups. These tools give admins

more clarity about how and when we enforce our policies in their Groups and gives them greater visibility into what is happening in their communities. This also means that they're more accountable for what happens under their watch. We help admins to establish positive Group norms by adding a section for rules so they can be clear about what is and isn't allowed. Admins and moderators also have the option to share which rule a member broke when declining a pending post, removing a comment, or muting a member.

Finally, we reach out to law enforcement whenever we see a credible threat of imminent harm. We contact federal, state, or local law enforcement depending on the specific circumstances of a threat. We have a long history of working successfully with the Department of Justice, the FBI, state and local law enforcement, and other government agencies to address a wide variety of threats to our platform. We have been able to provide support to authorities around the world. We have strict processes in place to handle government requests we receive, and we disclose account records in accordance with our terms of service and applicable law. We also have law enforcement response teams available around the clock to respond to emergency requests.

Questions from Representative Himes

For Facebook

1. Please provide the analyses that back the statement that Facebook's users are not particularly drawn to clickbait and inflammatory content.

We know that one of the biggest issues social networks face is that, when left unchecked, people will engage disproportionately with more sensationalist and provocative content. At scale such content can undermine the quality of public discourse and lead to polarization. In our case, it can also degrade the quality of our services. Our research suggests that no matter where we draw the line for what is allowed, as a piece of content gets close to that line, people will engage with it more on average—even when they tell us afterwards they don't like the content.

That is why we've invested heavily in our integrity teams—now totaling about 35,000 people—and have taken steps to minimize the amount of divisive news content people see in News Feed, including by reducing clickbait headlines.

Clickbait intentionally omits crucial information or exaggerates the details of a story to make it seem like a bigger deal than it really is. One of our News Feed values is authentic communication. We've heard from people that they prefer to see clearly written headlines that help them decide how they want to spend their time, and that authentic stories are the ones that resonate most—those that people consider genuine and not misleading, sensational, or spammy. Moreover, one of the signals we use to assess clickbait, beyond the specific wording of the headline, is whether users who click on these links tend to quickly come back to Facebook, suggesting disappointment with what people found on the landing page.

2. Please provide any analysis documenting how Facebook users are actually behaving as a result of Facebook's algorithm, and please include a description of specific changes made to the algorithms that have resulted in either 1) reduced presentation of clickbait or inflammatory content to users, or 2) reduced engagement with clickbait or inflammatory content by users.

We value authentic communication on our platform because people have told us they like seeing authentic stories the most. That's why we work hard to understand what type of stories and posts people consider genuine, so we can show more of them in News Feed. We also work to understand what kinds of stories people find misleading and spammy to help make sure people see those less. That includes clickbait headlines that are designed to get attention and lure visitors into clicking on a link. In an effort to support an informed community, we're always working to determine what stories might have clickbait headlines so we can show them less often.

We've made a variety of changes to News Feed to reduce the distribution of stories from sources that consistently post clickbait headlines that withhold and exaggerate information. For example, we reduce the distribution of posts that lead people to click and then quickly come back to News Feed. We also built an automated system that uses a set of identified clickbait headlines to determine phrases that are commonly used in clickbait and not in other headlines, so that we can identify and reduce the distribution of clickbait articles.

Questions from Representative Carson

For all witnesses

1. Can you provide a brief update on the policies that your companies currently use to address the threat deepfakes or other sophisticated manipulated media pose to users? What is your current approach, and how confident are you that you can identify and stop a foreign-connected deepfake as part of an attempted online influence operation?

This year, we announced our policy toward misleading manipulated videos. We remove misleading manipulated video if it meets the following criteria:

(1) It has been edited or synthesized—beyond adjustments for clarity or quality—in ways that aren't apparent to an average person and would likely mislead someone into thinking that a subject of the video said words that they did not actually say; and

(2) It is the product of artificial intelligence or machine learning that merges, replaces, or superimposes content onto a video, making it appear to be authentic.

This policy does not extend to content that is parody or satire, or video that has been edited solely to omit or change the order of words.

Consistent with our existing policies, audio, photos, or videos—whether a deepfake or not—will be removed from Facebook if they violate any of our other Community Standards, including those governing nudity, graphic violence, voter suppression, and hate speech. And videos that don't meet these standards for removal are still eligible for review by one of our independent third-party fact-checkers, which include over seventy partners worldwide fact-checking in over fifty languages. If content is rated false or partly false by a fact-checker, we significantly reduce its distribution in News Feed and reject it if it's being run as an ad. And critically, people who see it, try to share it, or have already shared it will see warnings alerting them that it's been rated false by a fact-checker.

Our enforcement approach has several components, from investigating AI-generated content and deceptive behaviors like fake accounts; to partnering with academia, government, and industry; to exposing people behind these efforts.

We also continue to invest in partnerships, including with journalists, academics, and independent fact-checkers, to help us reduce the distribution of false news and misinformation, as well as to better inform people about the content they encounter online. Last year, we launched the Deepfake Detection Challenge, which spurred people from all over the world to produce more research and open source tools to detect deepfakes. This project, supported by $10 million in grants, included a cross-sector coalition of organizations in civil society and the technology, media, and academic communities. We recently announced the results of the challenge and will be continuing to work with the technical communities to advance solutions in this space. For more information, please visit https://ai.facebook.com/blog/deepfake-detection-challenge-results-an-open-initiative-to-advance-ai.

In a separate effort, we've also partnered with Reuters, the world's largest multimedia news provider, to help newsrooms worldwide to identify deepfakes and manipulated media through a free online training course. News organizations increasingly rely on third parties for large volumes of images and video, and identifying manipulated visuals is a significant challenge. This program aims to support newsrooms trying to do this work.

2. I know that there was reporting in December about accounts associated with the Epoch Times media outlet as having used faked profile photos on Facebook. Has Facebook, or the other companies, identified any new deployments of deepfakes in a fashion such as this, particularly if linked to a state actor?

Facebook is constantly monitoring for inauthentic behavior on its platform. When it comes to deepfakes, we know the technology is advancing and we are focused on getting ahead of the threat. However, it is important to note that our security work against influence operations, like the one we removed in December 2019, is focused on identifying the patterns of misleading behavior which allows us to consistently detect violating behaviors and remove them from our platforms. In this case, we identified the network of accounts and Pages based on their behavior, including the use of fake accounts at the core of the operation, regardless of whether they relied on AI-generated profile photos.

We're constantly working to find and stop coordinated campaigns that seek to manipulate public debate across our apps. In 2019 alone, we took down over fifty networks for engaging in coordinated inauthentic behavior, including ahead of major democratic elections. So far this year, we've taken down twenty-two networks that were engaging in this sort of deceptive behavior. If we find instances of coordinated inauthentic behavior conducted on behalf of a government entity or by a foreign actor, we apply the broadest enforcement measures, including the removal of every on-platform property connected to the operation itself and the people and organizations behind it.

We share information about all of the coordinated inauthentic networks we take down every month. For more information, please see https://about.fb.com/news/tag/coordinated-inauthentic-behavior/.

3. Throughout the recent protests in the wake of George Floyd's murder, some white nationalist groups have pushed messages of hate and violence, in an attempt to undermine the legitimacy of the protest movement. One such white nationalist group, Identity Evropa, actually created a fake Twitter account, impersonated a left-wing Antifa activist, and explicitly called for violence during some of the most tense moments of the protests. With this example in mind:

a. How do your companies assess and evaluate any attempts by foreign actors to manipulate the information environment or create chaos during such fast-moving and emotional charged events? Especially when weighed against social media's role as an engine for legitimate civic organizing and the airing of genuine political or social grievances, as we've seen nationwide this month?

If we find instances of coordinated inauthentic behavior—whether domestic or foreign—we remove both the accounts and the content connected to the operation itself and the people and organizations behind it.

We recognize these are incredibly difficult and challenging times, and that's why it is more important than ever that people can have authentic conversations on our platforms about issues that matter to them, including social and racial injustice. We also know, however, that malicious actors are working to interfere with and manipulate these conversations, undermine the effectiveness of our public health responses, encourage social unrest, promote fraud, influence our elections, and make inauthentic behavior seem genuine. Stopping these bad actors is one of our highest priorities, and we continue to work tirelessly to do so.

Specifically on the US protests, since the protests started, we've seen some speculation around concerns about coordinated inauthentic behavior or foreign interference. We have been actively looking and we haven't yet seen foreign interference or domestic coordinated inauthentic behavior targeting the protests. As many in the research community and industry have done, we want to caution people from jumping to conclusions without clear evidence of foreign interference. We know that one goal of influence operations is to make their perpetrators look more powerful than they are. Speculation like this plays right into the hands of these bad actors—it can make us distrust each other and delegitimize authentic advocacy and political organizing, essential pillars of democracy.

In addition to our coordinated inauthentic behavior work, we also look for and enforce against low sophistication inauthenticity on our platforms. As your question references, there has been some public reporting about inauthentic actors trying to infiltrate Antifa groups. We've seen some examples of isolated efforts by people in the US using multiple accounts to pose as Antifa members. We've taken down those fake accounts. For example, we took down a handful of largely dormant Pages and accounts connected to the account posing as Antifa that Twitter linked to white supremacists. We haven't seen them post the same content that had been posted on Twitter. We continue monitoring and will take action as we find violations.

b. Can your company provide an update on the procedures that it currently uses to identify content that incites violence? Are those processes automated, or how does that process currently work? What definitions are used, since I imagine the it's not always clear-cut?

We aim to prevent potential offline harm that may be related to content on Facebook. Accordingly, we remove language that incites or facilitates serious violence. We also ban groups that proclaim a hateful and violent mission from having a presence on our apps and we remove content that represents, praises, or supports them. To date, we've identified a wide range of groups across the globe as hate organizations because they engage in coordinated violence against others based on characteristics such as religion, race, ethnicity, or national origin, and we routinely evaluate groups and individuals to determine if they violate our policy. In fact, last month we designated and banned from our platform a violent, US-based anti-government network because it actively seeks to commit violence.

Moving fast to find and remove dangerous organizations, including terrorist and hate groups, takes significant investment in both people and technology. At Facebook, we have 350 people who exclusively or primarily focus on countering dangerous organizations as their core responsibility. This group includes former academics who are experts on counterterrorism, former prosecutors and law enforcement agents, investigators and analysts, and engineers. We also have tripled the size of our teams working in safety and security since 2016 to over 35,000 people—including teams that review reports of hate speech and content that praises, supports, or represents hate groups twenty-four hours a day, seven days a week.

Three years ago, we started to develop a playbook and a series of automated techniques to detect content related to terrorist organizations such as ISIS, al Qaeda, and their affiliates. We've since expanded these techniques to detect and remove content related to other terrorist and hate groups. We're now able to detect text embedded in images and videos in order to understand its full context, and we've built media matching technology to find content that's identical or near-identical to photos, videos, text, and even audio that we've already removed. When we started detecting hate organizations, we focused on groups that posed the greatest threat of violence at that time, and we've now expanded to detect more groups tied to different hate-based and violent extremist ideologies and using different languages. In addition to building new tools, we've also adapted strategies from our counterterrorism work, such as leveraging off-platform signals to identify dangerous content on Facebook, and implementing procedures to audit the accuracy of our AI's decisions over time.

In the first three months of 2020, we removed about 4.7 million pieces of content on Facebook connected to organized hate—which encompasses a range of groups across the globe because they engage in coordinated violence against others based on characteristics such as religion, race, ethnicity, or national origin. We routinely evaluate groups and individuals to determine if they violate our policy. The amount of content we removed in Quarter 1 of 2020 was an increase of over 3 million pieces of content from the previous quarter. Additionally, we increased our proactive detection rate for organized hate, or the percentage of content we remove that we detect before someone reports it to us, from 89.6% in Quarter 4 of 2019 to 96.7% in Quarter 1 of 2020. We saw similar progress on Instagram where our proactive detection rate increased from 57.6% to 68.9%, and we removed 175,000 pieces of content in Quarter 1 of 2020, up from 139,800 the previous quarter.

In addition, since we built this system for content tied to hate groups based on what we learned from detecting terrorist content, we've been able to identify where content related to one problem is distinct from the other. For example, we've seen that violations for organized hate are more likely to involve memes, while terrorist propaganda is often dispersed from a central media arm of the organization and includes formalized branding. Identifying these patterns helps us continue to fine tune the systems for detecting organized hate and terrorist content.

For Facebook

1. **Back in November 2017, I asked Mr. Stretch of Facebook about identifying Russian- backed, inauthentic accounts on the platform and how those messages were able to maintain their presence on the platform throughout, and even after the 2016 election. Those Russian-backed accounts were pushing narratives of some of**

the most divisive identity politics in the United States, to sow division among Americans.

a. Can you provide an update on the guidelines Facebook and Instagram utilize for preventing malign disinformation in Ad Content?

We investigate and enforce against any type of inauthentic behavior. If we find instances of coordinated inauthentic behavior conducted on behalf of a government entity or by a foreign actor, we remove both the accounts and the content connected to the operation itself and the people and organizations behind it.

Regarding ads posted by authentic accounts, ads must comply with our Advertising Policies, which can be found at https://www.facebook.com/policies/ads. Our Advertising Policies prohibit ads that include claims debunked by third-party fact-checkers or, in certain circumstances, claims debunked by organizations with particular expertise. Advertisers that repeatedly post information deemed to be false may have restrictions placed on their ability to advertise on Facebook.

Ads are subject to Facebook's ad review system, which relies primarily on automated tools to check ads against these policies. We use human reviewers to improve and train our automated systems and, in some cases, to review specific ads. This review happens before ads begin delivering, but may also happen after, if people hide, block, or provide negative feedback about an ad. When we detect an ad that violates our Advertising Policies, we disapprove it.

b. How were policies and enforcement mechanisms "fine-tuned" since 2017 to better detect attempts at malign disinformation through paid ads?

We're constantly working to find and stop coordinated campaigns that seek to manipulate public debate across our apps. In 2019 alone, we took down over fifty networks worldwide for engaging in coordinated inauthentic behavior, including ahead of major democratic elections. And to date, we have taken down twenty-two networks that were engaging in this sort of deceptive behavior, including three networks originating from Russia, two from Iran, and three based here in the United States.

Over the past four years, we've built a global team of more than 35,000 people working across the company on issues to secure the safety and security of our services, including combating coordinated inauthentic behavior. In December of 2016, we launched an independent fact-checking program that became the basis for evaluating whether content posted to Facebook is actually true. Content found to be false or partly false by our fact-checking partners is labelled via an overlaid warning screen and its distribution is reduced. We've since expanded the program to Instagram and now have more than seventy fact-checking partners covering more than fifty languages around the world.

c. Do you feel confident that your policies and tools are tight enough to keep malign foreign elements from leveraging Facebook or Instagram ads in influence operations? Or are there scenarios of concern, and what are you doing to stay ahead of it?

As described in response to your Question 1(b), we're constantly working to find and stop coordinated campaigns that seek to manipulate public debate across our apps.

As part of that effort, we will continue our work to detect malicious behavior and to enforce against violations of our Terms and Advertising Policies. We are making progress rooting out this abuse, but it's an ongoing effort. We're committed to continually improving to stay ahead. That means building better technology, hiring more people, and working more closely with law enforcement, security experts, and other companies.

Questions from Representative Swalwell

For all witnesses

1. Do your platforms have a policy to combat anti-vaccine misinformation in posts by users? Does that policy extend beyond demonetization, if relevant? If so, how?

Because misinformation about health topics can be especially problematic, we commenced an effort to supplement the work of our third-party fact-checkers for misinformation about vaccinations. Specifically, we rely on the publicly available work of leading health organizations on the issue of vaccines, such as the US Centers for Disease Control and Prevention ("CDC") and the World Health Organization ("WHO"), among others, to identify verifiable hoaxes on the topic. An example of a claim that has been widely disproven by these organizations is the assertion that vaccines cause autism.

We take a number of different steps to substantially reduce the distribution of these publicly identified vaccine hoaxes across our platform. First, if an ad includes this type of misinformation about vaccinations, it will be rejected. Beyond advertisements, when we become aware of Groups or Pages on Facebook that propagate this type of misinformation, we remove them from recommendation surfaces on the platform and from predictions when you type into search. We likewise won't show or recommend content that contains this misinformation about vaccinations on Instagram Explore or hashtag Pages. Furthermore, all content from offending Groups and Pages will be demoted in News Feed using our ranking systems, and the Groups and Pages themselves will be demoted in search results. When Pages repeatedly post misinformation about vaccinations, they will lose access to our fundraising tools.

Because vaccine hoaxes have been previously (and publicly) debunked by expert health organizations, politicians that post this content would be treated the same as all other users—their organic content would be downranked and their ads would be rejected.

Consistent with our overall approach to combating misleading or false information, in addition to reducing its distribution, we seek to inform users with additional context on the topic. For vaccinations, we have gone further and launched educational modules that pop up for US-based users when they engage with content about vaccines, including but not limited to misinformation about vaccines. The educational modules appear on Instagram as well as in Facebook search, invitations to join Groups, and on Pages. The modules provide US users with authoritative context and other resources from the CDC.

We are working to apply the steps we are taking to combat misinformation about vaccinations to misinformation about other important health topics.

2. Do your platforms have a policy to combat public health misinformation in posts by users? Does that policy extend beyond demonetization, if relevant? If so, how?

Facebook is designed to give people a voice, and we encourage a wide array of expression on our platform. At the same time, we have an important role to play in keeping abuse off our platform, especially when it comes to advertisements, and we are committed to

making Facebook a safe place for users. Health content, given its often sensitive nature, features prominently in that commitment.

Facebook is dedicated to reducing the spread of misinformation on our platform. We use multiple means to achieve that goal, including removing fake accounts, disrupting the financial incentives behind propagating false and misleading information, working with third-party fact-checkers to let people know when they are reading or sharing information that has been disputed or debunked, and limiting the distribution of stories that have been flagged as false or misleading by these fact-checkers. The third-party fact-checkers with which we work—who are signatories to the non-partisan International Fact-Checking Network Code of Principles—investigate claims and make determinations about a post's truth or falsity. This fact-checking process also applies to misinformation about health. In addition, as explained in response to your Question 1, we supplement the work of our third-party fact-checking partners through our approach to misinformation about vaccinations, and we are exploring ways to apply that approach to other areas of public health.

In addition, when it comes to misinformation about COVID-19 specifically, we remove COVID-19 related misinformation that could contribute to imminent physical harm, such as posts that make false claims about cures, treatments, the availability of essential services, or the location and severity of the outbreak. We regularly update the claims that we remove based on guidance from the WHO and other health authorities. For claims that don't directly result in physical harm, like conspiracy theories about the origin of the virus, we continue to work with our network of third-party fact-checkers, as described above. During the month of April, we put warning labels on about 50 million pieces of content related to COVID-19 on Facebook, based on around 7,500 articles by our independent fact-checking partners.

Ever since COVID-19 was declared a global public health emergency in January, we've been working to connect people to accurate information from health experts and keep harmful misinformation about COVID-19 from spreading on our apps. We've now directed over 2 billion people to resources from the WHO and other health authorities through our COVID-19 Information Center and pop-ups on Facebook and Instagram, with over 350 million people clicking through to learn more.

3. Has One American News Network (OANN) had videos or posts removed from your platform? If so, how many and for what reasons?

While we do not typically comment on specific cases of content removal for privacy reasons, when we identify or learn of content that violates our policies, we remove that content regardless of who posted it. Decisions about whether to remove content are based on our Community Standards. The political affiliation of the user generating the content has no bearing on that content removal assessment. We have removed content posted by individuals and entities across the political spectrum.

4. Has Fox News had videos or posts taken removed from your platform? If so, how many and for what reasons?

While we do not typically comment on specific cases of content removal for privacy reasons, when we identify or learn of content that violates our policies, we remove that content regardless of who posted it. Decisions about whether to remove content are based on our Community Standards. The political affiliation of the user generating the content has no bearing on that content removal assessment. We have removed content posted by individuals and entities across the political spectrum.

5. Has The Epoch Times had videos or posts removed from your platform? If so, how many and for what reasons?

While we do not typically comment on specific cases of content removal for privacy reasons, in this case we've released some information, given evidence of coordinated inauthentic behavior. In 2018, Facebook prohibited NTD Television, an Epoch Times sister company, from advertising on the platform due to a high volume of violations of our ads policies, including low quality or disruptive content and sensational content, and circumvention of our systems. In August 2019, we subsequently prohibited Epoch Media Group from advertising on the platform due to further violations of our ads policies, including those regarding political ads transparency.

In December 2019, Facebook also removed a network of coordinated inauthentic behavior that originated in Vietnam and the US and was focused primarily on the US and some Vietnamese, Spanish, and Chinese-speaking audiences globally. This activity primarily focused on The BL, a US-based media company, and its Pages, which were operated by individuals in the US and Vietnam. The people behind this activity made widespread use of fake accounts—many of which had been automatically removed by our systems—to manage Pages and Groups, and to automate posting at very high frequencies and direct traffic to off-platform sites. Some of these accounts used profile photos that were generated by artificial intelligence and masqueraded as Americans in order to join Groups and post The BL content. To evade our enforcement, they used a combination of fake and authentic accounts of local individuals in the US to manage Pages and Groups. Although the people behind this network attempted to conceal their identities and coordination, our investigation linked this activity to Epoch Media Group and individuals in Vietnam working on its behalf. The BL-focused network repeatedly violated a number of our policies, including our policies against coordinated inauthentic behavior, spam, and misrepresentation, to name just a few. The BL is now banned from Facebook. We are continuing to investigate all linked networks, and will take action as appropriate if we determine they are engaged in deceptive behavior.

6. On June 18, 2020, Facebook, Instagram, and Twitter removed a Trump campaign ad featuring a symbol (a red inverted triangle) used by Nazis to designate political prisoners in concentration camps. Facebook, which owns Instagram, stated, "We removed these posts and ads for violating our policy against organized hate. Our policy prohibits using a banned hate group's symbol to identify political prisoners without the context that condemns or discusses the symbol."

a. How many symbols of hate would a campaign or candidate have to run before the campaign's account or page would be taken down from your platform?

b. How many false or partly false posts, videos, or ads would a campaign or candidate have to run before the campaign or candidate's account or page would be taken down from your platform? Or would consistent posting of false or partly false posts or ads go unenforced?

c. Have campaign or candidate accounts, pages, or channels associated with U.S. persons been taken down because of repeated posting or advertising of false or partly false information? If so, how many? And if not, have you taken other actions against said accounts, pages, or channels?

d. Have campaign or candidate accounts, pages, or channels associated with U.S. persons been taken down because of repeated use – whether through advertising or not – of symbols of hate and/or violating anti-hate policies? If so, how many? And if not, have you taken other actions against said accounts, pages, or channels?

e. Are your platforms considering implementing new policies or revising existing ones to address the issues raised in questions 7a through 7d?

Under our Community Standards, we do not allow the use of symbols that represent organizations or individuals involved in organized hate to be shared on our platform without context that condemns or neutrally discusses the content. We also prohibit hate speech, bullying, intimidation, and other kinds of abusive behavior. This is true for all organic content and ads.

While we do not typically comment on specific cases of content removal for privacy reasons, when we identify or learn of content that violates our policies, we remove that content regardless of who posted it. Decisions about whether to remove content are based on our Community Standards and Advertising Policies.

We don't want people to game the system, so we do not share the specific number of violations that leads to a temporary block or permanent suspension. When we remove content for violating our policies, we notify the person who posted it to explain why, with some narrow exceptions to account for things like child exploitation imagery.

If someone violates our policies multiple times, their account will be temporarily blocked; a Page that does so will be unpublished. When a person is in a temporary block, they can read things on Facebook, but they can't like, comment, or post. If that person is also the admin of a Facebook Page, the block prevents them from posting to the Page. Pages that repeatedly violate our policies may also lose the ability to advertise on Facebook.

We also work to keep confirmed misinformation from spreading. For example, we reduce its distribution in News Feed so fewer people see it. And if Pages, domains, or Groups repeatedly share misinformation, we'll continue to reduce their overall distribution, and we'll place restrictions on the Pages' ability to advertise and monetize.

We regularly review our policies to make sure they are in the right place.

Questions from Rep. Maloney

For all witnesses

1. Recognizing that strides have been made since 2016 through 2018:

a) Is it your company's stance that that the current volume and types of indicators, data, and/or metadata about potential foreign influence activity shared both within the industry and between the industry and the U.S. government are sufficient for protecting our national conversation and elections from foreign influence or interference moving forward?

Inauthentic behavior, including foreign influence campaigns, has no place on Facebook. If we find any instances of coordinated inauthentic behavior targeting the US conducted on behalf of a foreign actor, we apply the broadest enforcement measures, including the removal of every on-platform property connected to the operation itself and the people and organizations behind it. To date, we have taken down twenty-two networks that were engaging in this sort of deceptive behavior, including three networks originating from Russia, two from Iran, and three based here in the United States.

We know that inauthentic behavior is not limited to a specific type of technology or service. The better we can be at working together with industry and outside security researchers, the better we'll do by our community. We continuously look for ways to enhance our collaboration with industry and the security research community while ensuring that we put the right checks in place to protect people's information.

That's why we're working closely with our fellow tech companies to deal with the threats of inauthentic behavior we have all seen. Several takedowns that we conducted and announced were in close collaboration with other tech platforms, security companies, and law enforcement agencies. For instance, in March 2020, we took down a network of accounts engaging in foreign interference in Ghana and Nigeria on behalf of Russia targeting primarily the United States. We shared this information with other tech companies, including Twitter, which also announced the takedown of this activity on its platform. For more information, see https://about.fb.com/news/2020/03/removing-coordinated-inauthentic-behavior-from-russia/.

We are also committed to working with law enforcement, and we deeply respect and support the work law enforcement agencies do to keep us safe. We have a long history of working successfully with law enforcement, including the FBI and DHS, to address foreign and domestic influence operations.

b) What limits imposed by U.S. law or regulations might prevent your compnay from maximally sharing data or metadata associated with high-confidence foreign influence operations/CIB with U.S. law enforcement?

As Nathaniel Gleicher testified, information sharing among the industry and the government has improved over the past few years. And we work closely with law enforcement, industry partners, and civil society. That said, the industry would benefit from a clear legal framework regarding data sharing in the context of influence operations.

We continuously look for ways to enhance our collaboration with industry and the security research community while ensuring that we put the right checks in place to protect people's information, because we know that inauthentic behavior is not limited to a specific type of technology or service. The better we can be at working together with industry and outside security researchers, the better we'll do by our community.

c) How might relevant changes to the Secure Communications Act (SCA), the Electronic Communications Privacy Act (ECPA), Cybersecurity Information Sharing Act (CISA), or the Section 230 Communications Decency Act (CDA) help or harm your companies' efforts to prevent foreign influence from infiltrating your platforms?

Section 230 of the Communications Decency Act has been essential to protecting free expression and innovation on the internet, and we believe its provisions are consistent with operating safe products that give consumers choice. Specifically in regard to fighting against foreign interference and inauthentic behavior, Section 230 allows us to do the work we need to do to keep people safe by enforcing our Community Standards.

d) Would considerations such as creating a "safe harbor" provision, or clearly delineating that assessed foreign influence actors don't have claim to the same data privacy protections as genuine users, affect those stances?

We support clarifying the role of platforms in combating influence operations, and we are always happy to discuss any legislation Members propose or offer in this space.

2. Would your company find valuable an Information Sharing and Analysis Center (ISAC) or equivalent formalized mechanism devoted specifically to data-sharing about potential foreign-linked influence operations? Would your company support a leading role in an ISAC or equivalent? Why or why not?

Facebook, along with industry partners, engages in a regular cadence of meetings with each other and with our government partners to discuss influence operation threats, trends, and developments, and cross-sector efforts to combat those through collective defenses. These meetings build on cross-industry and law enforcement information sharing, conducted in accordance with legal requirements.

When it comes to addressing the wide variety of threats to people on our platform, we have a strong and established track record of working together successfully with law enforcement, including the FBI and DHS, to address foreign and domestic influence operations.

We also share our insights on these operations via monthly transparency reports to ensure that what we are seeing and tracking is brought to a broader domain of public awareness, which in turn fosters a more resilient and secure information environment.

Questions from Rep. Krishnamoorthi

For Facebook

1. Mr. Gleicher referred to bans against Mr. Prigozhin and the IRA on the platform, but earlier he referred to removals against two hate groups themselves had previously been banned.

a) What do those bans specifically entail?

We ban groups that proclaim a hateful and violent mission from having a presence on our platform and we remove content that represents, praises, or supports them.

Facebook recently removed two networks of accounts for organizations that we consider to be hate groups: the Proud Boys and the American Guard, which we had previously banned from our platform but had attempted to return. While we had been investigating these two organizations' attempts to return for several weeks, we accelerated our investigation and enforcement after finding evidence that accounts from both organizations were discussing intentions to bring weapons to civil rights protests. In total, we removed 358 Facebook accounts and 172 Instagram accounts tied to the organization known as the Proud Boys. We removed 406 Facebook accounts and 164 Instagram accounts tied to the group known as the American Guard.

It's important to note that this is only a small piece of our expansive work to designate and ban hate groups from Facebook and continually enforce our policy so that we make Facebook as inhospitable to these groups as possible. While hate and terrorist groups are global issues, there is no one agreed upon standard for how to define them. So, in consultation with experts, we've developed our definitions to guide our designation process. As result, we've banned over 250 white supremacist organizations having determined they meet our criteria to be considered either a hate group or terrorist group and we continue to follow their attempts to return to the platform as well as remove content that praises, supports, or represents these organizations.

b) How specifically are actor bans enforced? Do they apply to both Instagram and Facebook?

Moving fast to find and remove dangerous organizations, including terrorist and hate groups, takes significant investment in both people and technology. At Facebook, we have 350 people who exclusively or primarily focus on countering dangerous organizations as their core responsibility. This group includes former academics who are experts on counterterrorism, former prosecutors and law enforcement agents, investigators and analysts, and engineers. We also have tripled the size of our teams working in safety and security since 2016 to over 35,000 people—including teams that review reports of hate speech and content that praises, supports, or represents hate groups twenty-four hours a day, seven days a week.

Three years ago, we started to develop a playbook and a series of automated techniques to detect content related to terrorist organizations such as ISIS, al Qaeda, and their affiliates. We've since expanded these techniques to detect and remove content related to other terrorist and hate groups. We're now able to detect text embedded in images and videos in order to

understand its full context, and we've built media matching technology to find content that's identical or near-identical to photos, videos, text, and even audio that we've already removed. When we started detecting hate organizations, we focused on groups that posed the greatest threat of violence at that time, and we've now expanded to detect more groups tied to different hate-based and violent extremist ideologies and using different languages. In addition to building new tools, we've also adapted strategies from our counterterrorism work, such as leveraging off-platform signals to identify dangerous content on Facebook, and implementing procedures to audit the accuracy of our AI's decisions over time.

In the first three months of 2020, we removed about 4.7 million pieces of content on Facebook connected to organized hate—which encompasses a range of groups across the globe because they engage in coordinated violence against others based on characteristics such as religion, race, ethnicity, or national origin. We routinely evaluate groups and individuals to determine if they violate our policy. The amount of content we removed in Quarter 1 of 2020 was an increase of over 3 million pieces of content from the previous quarter. Additionally, we increased our proactive detection rate for organized hate, or the percentage of content we remove that we detect before someone reports it to us, from 89.6% in Quarter 4 of 2019 to 96.7% in Quarter 1 of 2020. We saw similar progress on Instagram where our proactive detection rate increased from 57.6% to 68.9%, and we removed 175,000 pieces of content in Quarter 1 of 2020, up from 139,800 the previous quarter.

In addition, since we built this system for content tied to hate groups based on what we learned from detecting terrorist content, we've been able to identify where content related to one problem is distinct from the other. For example, we've seen that violations for organized hate are more likely to involve memes while terrorist propaganda is often dispersed from a central media arm of the organization and includes formalized branding. Identifying these patterns helps us continue to fine tune the systems for detecting organized hate and terrorist content.

c) Does the fact that previously banned entities were able to apparently reenter the platform mean that there are gaps Facebook still needs to address? And how is Facebook addressing those gaps?

We are committed to preventing terrorists and hate organizations from using Facebook, but just like offline, this work will never be completely finished and so we continue our efforts. We work to consistently enforce our policies, but we know that bad actors will try to come back to the platform. We remain vigilant in learning and combating new ways people may try to abuse our apps. We work with external partners to get the latest intelligence about adversarial behavior across the internet, and we commission independent research from academics and experts. We also learn from different teams at Facebook about successful methods in combating other forms of abuse that can be applied to this work.

For example, over the last six months, we worked with colleagues on our Threat Intelligence team to leverage their strategy for combating coordinated inauthentic behavior in order to develop a new tactic that targets a banned group's presence across our apps. We do this by identifying signals that indicate a banned organization has a presence, and then proactively investigating associated accounts, Pages, and Groups before removing them all at once. Once we remove their presence, we work to identify attempts by the group to come back on our platform.

We're also studying how dangerous organizations initially bypassed our detection, as well as how they attempt to return to Facebook after we remove their accounts, in order to strengthen our enforcement and create new barriers to keep them off our apps. We'll continue working to disrupt and remove dangerous organizations from our platform and we'll share how we're doing at enforcing our policies and combating new ways people may try to abuse our apps.

2. How would Facebook approach a scenario whereby a foreign government used its own Facebook/Instagram accounts (not state-controlled media) to overtly boost the President's controversial "looting starts, shooting starts" comments through advertising?

We do not permit advertisers to place ads about social issues, elections, or politics unless they confirm their ID as being in the country where they want to place such ads. We also require them to disclose who is responsible for the ad, which will appear on the ad itself. The ad and "Paid for by" disclaimer are placed in the Ad Library for seven years, along with more information such as range of spend and impressions, as well as demographics of who saw the ad. We already require that advertisers get authorized and add disclaimers to these ads in over fifty countries and territories, and now we're expanding proactive enforcement on these ads to countries where elections or regulations are approaching.

Advertisers on Facebook must comply with Facebook's Advertising Policies, including acknowledging that they are responsible for understanding and complying with all applicable laws and regulations. Therefore, violating the Federal Election Campaign Act also violates our terms.

It is hard for us to comment on hypotheticals, as we review each ad individually and proactively against our ad policies, which are, on the whole, stricter than our Community Standards. We do not have any newsworthy exceptions to our advertising policies and all advertisers are required to follow them.

We want to make sure the content people are seeing on Facebook is authentic. We believe that authenticity creates a better environment for sharing, and that's why we don't want people using Facebook to misrepresent who they are or what they're doing. We are committed to making Facebook a safe place. Expression that threatens people has the potential to intimidate, exclude, or silence others and isn't allowed on Facebook.