HEARING BEFORE
THE UNITED STATES HOUSE OF REPRESENTATIVES
COMMITTEE ON ENERGY & COMMERCE
SUBCOMMITTEE ON CONSUMER PROTECTION & COMMERCE

January 8, 2020

Testimony of Monika Bickert
Vice President for Global Policy Management, Facebook

I. Introduction

Chairwoman Schakowsky, Ranking Member McMorris Rodgers, and distinguished members of the Subcommittee, thank you for the opportunity to appear before you today. My name is Monika Bickert, and I am the Vice President of Global Policy Management at Facebook. In that role, I lead our efforts related to content policy and counterterrorism. Prior to assuming my current role, I served as lead security counsel for Facebook, working on issues ranging from children's safety to cybersecurity. And before that, I was a criminal prosecutor with the Department of Justice for 11 years in Chicago and Washington, DC, where I prosecuted federal crimes, including public corruption and gang violence.

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

II. Combating Manipulation and Deception

Community Standards

We publish Community Standards governing the types of content and behaviors that are acceptable on Facebook. For example, we prohibit hate speech, harassment, content posted by fake accounts, and—under a new policy—misleading manipulated media, including certain types of deepfakes. When we become aware of content that violates our Community Standards, through either proactive technical measures or reports, we remove it.

Some types of misinformation—such as attempts to interfere with or suppress voting or participation in the census—violate our Community Standards, and we work proactively to remove this type of harmful content. We are mindful of our responsibility to respect freedom of expression, but our Community Standards are clear that we remove content when it has the potential to contribute to offline physical harm.

We recognize the risks of manipulated media. Manipulated media can be made with simple technology like Photoshop, or with sophisticated tools that use artificial intelligence or "deep learning" techniques to create videos that distort reality—usually called "deepfakes." While these videos are still relatively rare on the internet, they present a significant challenge for our industry and society as their use increases, and we have been engaging broadly with internal and external stakeholders to better understand and address this issue. Based on these conversations, we have been considering a number of options regarding misleading manipulated media, including deepfakes. That is why we just announced a new policy that we will remove certain types of misleading manipulated media from our platform. In particular, under this policy, which is part of our Community Standards, we will remove videos that have been edited or synthesized using artificial intelligence or deep learning techniques in ways that are not apparent to an average person and that would mislead an average person to believe that a subject of the video said words that they did not say. The policy is designed to prohibit the most sophisticated attempts to mislead people.

To be clear, forms of misleading manipulated media that do not meet these criteria—such as videos that have been edited solely by splicing to omit or change the order of words, or parodies or satires—are still subject to our other Community Standards and are eligible for fact-checking. For example, a synthesized video of a celebrity in which the celebrity is nude would violate our nudity policies. Manipulated media may also be spread in a coordinated manner by fake accounts, which would violate our policies against inauthentic behavior; in such cases, the content posted by such accounts would also be removed.

Misinformation

We recognize that some types of misleading information lack quality and integrity, despite not directly violating our Community Standards. Our approach to such misinformation has several components, including working with independent, third-party fact-checkers to help reduce the spread of false news and other types of viral misinformation; investigating AI-generated content and deceptive behaviors like fake accounts; partnering with academia, government, and industry on tackling broad issues; and exposing the bad actors behind these efforts.

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. Manipulated photos and videos can be fact-checked by one of our independent, third-party fact-checking partners, who are certified through the non-partisan International Fact-Checking Network. We now have over 50 partners around the world fact-checking content in over 40 languages, and we are investing in ways to scale these efforts further. Fact-checkers use their own expertise to determine which stories to review; many of our third-party fact-checking partners focus on misinformation in images and videos. This includes identifying when an image or video is being presented out of context using tools such as reverse image search or

utilizing video editing programs to identify when manipulation has occurred. Fact-checking partners are able to assess the truth or falsity of a photo or video by combining these skills with original reporting, including outreach to technical experts, academics, or government agencies.

Once a fact-checker rates a photo or video as false or partly false, we reduce its distribution in News Feed and reject it if it's being run as an ad. We also implement an overlaid warning screen on top of photos and videos marked as false. If people try to share the content, they will be notified of the additional reporting. They will also be notified if content they have shared in the past has since been rated false by a fact-checker.

Moreover, in order to more effectively fight false news, we also take action against Pages and domains that repeatedly share or publish content which is rated as false. Such Pages and domains will see their distribution reduced as the number of offenses increases. Their ability to monetize and advertise will be removed after repeated offenses. Over time, Pages and domains can restore their distribution and ability to monetize and advertise if they stop sharing false news.

We also use machine learning to assist in our fight against misinformation. Algorithms cannot fundamentally tell what content is true or false, but they do help in the process. For example, our machine learning models use various signals to identify content which might be false or partly false. Comments expressing disbelief are one signal that helps inform our prediction, as well as feedback from our community when people mark something as false news. And we use model predictions to prioritize the content we show third-party fact-checkers. Since there are hundreds of millions of pieces of content per week shared on Facebook, we prioritize third-party fact-checkers' time. In addition to helping us predict content for fact-checkers to review, machine learning helps us identify duplicates of debunked stories. In turn, fact-checker ratings help further train our machine learning model, so it's a cyclical process.

We are always improving our policies and enforcement practices, and we will continue to closely monitor this issue and to consult with external stakeholders to ensure we're taking the right approach. Across the world, we've been driving conversations with more than 50 global experts with technical, policy, media, legal, civic, and academic backgrounds to inform our policy development. As these partnerships and our own insights evolve, so too will our policies toward manipulated media. In the meantime, we're committed to investing within Facebook and to working with other stakeholders in this area to find solutions with real impact.

Coordinated Inauthentic Behavior

The idea behind Facebook is to help bring communities together in an authentic way. We believe that people are more accountable for their statements and actions when they use their authentic identities. Fake accounts are often behind harmful and misleading content, and we work hard to keep them off Facebook. We took down over 5 billion fake accounts

in the first three quarters of 2019, and our technology stopped millions of additional attempts every day to establish fake accounts before they were created. When we take down these accounts, it's because of their deceptive behavior (like using networks of fake accounts to conceal their identity); it's not based on the actors behind them or what they say.

Our efforts to prevent coordinated inauthentic behavior focus on four areas. First, our expert investigators use their experience and skills in areas like cybersecurity research, law enforcement, and investigative reporting to find and take down the most sophisticated threats. Second, we build technology to detect and automatically remove the most common threats. Third, we provide transparency and reporting tools so users can make informed choices when they encounter borderline content or content that we miss. We publicize our takedowns of coordinated inauthentic behavior for all to see, and we also provide information about them to third parties for their review and share relevant data with researchers, academics, and others. Fourth, we work closely with civil society, researchers, governments, and industry partners, so they can flag issues and we can work to resolve them quickly. Engaging with these partners regularly helps us improve the efficacy of our techniques and learn from their experiences.

Using this combination of tools, we continually adapt our platforms to make deceptive behaviors much more difficult and costly. When we conduct a takedown, we identify the tactics the bad actors used, and we build tools into our platforms to make those tactics more difficult at scale. By continuing to develop smarter technologies, enhance our defenses, improve transparency, and build strong partnerships, we are making the constant improvements we need to stay ahead of our adversaries and to protect the integrity of our platforms.

III. Partnering to Improve Deepfake Detection

Deepfake techniques have significant implications for determining the legitimacy of information presented online. Yet researchers in academia, government, and industry still lack strong data sets to analyze and benchmark this challenge. We want to encourage additional research and development in this area and ensure there are better open-source tools to detect deepfakes. That's why Facebook has partnered with a cross-sector coalition of organizations including the Partnership on AI, Cornell Tech, the University of California Berkeley, MIT, WITNESS, Microsoft, the BBC, and AWS, among several others in civil society and the technology, media, and academic communities to build the Deepfake Detection Challenge.

The goal of the Challenge is to produce technology that everyone can use to better detect when AI has been used to alter a video in order to mislead the viewer. The Deepfake Detection Challenge includes a data set and leaderboard, as well as grants and awards, to spur the industry to create new ways of detecting and preventing media manipulated via AI from being used to mislead others. The governance of the Challenge will be facilitated and overseen by the Partnership on AI's new Steering Committee on AI and Media Integrity, which is comprised of members including Facebook and others in civil society

and the technology, media, and academic communities.

It's important to have data that is freely available for the community to use. That is why we constructed a new training data set specifically for this Challenge, working with a third-party vendor to engage a diverse set of individuals who agreed to participate in creating the data set for the Challenge. We are also funding research collaborations and prizes for the Challenge to help encourage more participation. In total, we are dedicating significant resources to fund this industry-wide effort.

We also recently launched an e-learning course as a partnership between the Facebook Journalism Project and Reuters Institute. The course, titled "Identifying and Tackling Manipulated Media," aims to help newsrooms around the world equip themselves to identify manipulated media. It includes real-world examples, hypothetical cases, and insights into the evolving technology used to create and detect manipulated media, including deepfakes. It teaches journalists about the various types of altered media and the ways in which newsrooms can confidently verify and publish truthful content from third-party sources.

Manipulated media presents a constantly evolving challenge, and our hope is that by helping the industry and the AI community come together, we can make faster progress in a way that benefits the whole of society.

IV. Conclusion

We recognize that both the issues and challenges in addressing manipulated media are rapidly evolving. Experts have called on the industry to come together to develop a consistent approach across platforms. As they have pointed out, consistent enforcement across platforms is important to protect consumers from such content migrating from platform to platform. We agree. As our CEO Mark Zuckerberg has said, we need to develop consistent industry standards on issues such as manipulated media. We have encouraged the industry through our trade association to work together—specifically on manipulated media—in a more uniform way, pushing for common standards and a consistent approach across platforms. We welcome the opportunity to collaborate and partner with other industry participants and interested stakeholders, including academics, civil society, and lawmakers to help develop such an approach.

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

Thank you, and I look forward to your questions.