RPTR FORADORI

EDTR CRYSTAL

NATIONAL SECURITY CHALLENGES OF ARTIFICIAL INTELLIGENCE, MANIPULATED MEDIA, AND "DEEPFAKES"

Thursday, June 13, 2019

U.S. House of Representatives,

Permanent Select Committee on Intelligence,

Washington, D.C.

The committee met, pursuant to call, at 9:00 a.m., in Room 1100, Longworth House Office Building, the Honorable Adam Schiff (chairman of the committee) presiding.

Present: Representatives Schiff, Himes, Sewell, Quigley, Castro, Heck, Welch, Maloney, Demings, Krishnamoorthi, Nunes, Wenstrup, Stewart, Crawford, Hurd, and Ratcliffe.

The Chairman. The committee will come to order.

Before we begin, I want to remind all members that we are in open session, and as such, we will discuss unclassified matters only.

Please have a seat.

Our members may be wondering in a bit late. We were here until 1 in the morning, but those on Armed Services were here until about 5 or 6 in the morning. So we have a few groggy members here on the committee.

In the heat of the 2016 election, as the Russian hacking and dumping operation became apparent, my predominant concern was that the Russians would begin dumping forged documents along with the real ones that they stole. It would have been all too easy for Russia for another malicious actor to seed forged documents among the authentic ones in a way that would make it almost impossible to identify or rebut the fraudulent material. Even if a victim could ultimately expose the forgery for what it was, the damage would be done.

Three years later we are on the cusp of a technological revolution that could enable even more sinister forms of deception and disinformation by malign actors, foreign or domestic. Advances in AI and machine learning have led to the emergence of advanced digitally doctored types of media, so-called "deepfakes," that enable malicious actors to foment chaos, division, or crisis, and they have the capacity to disrupt entire campaigns, including that for the Presidency.

Rapid progress in artificial intelligence algorithms has made it possible to manipulate media -- video, imagery, audio, text -- with incredible, nearly imperceptible results. With sufficient training data, these powerful deepfake generating algorithms can portray a real person doing something they never did or saying words they never uttered.

These tools are readily available and accessible to both experts and novices alike, meaning that attribution of a deepfake to a specific author, whether a hostile intelligence service or a single internet troll, will be a constant challenge.

What is more, once someone views a deepfake or a fake video, the damage is largely done. Even if later convinced that what they have seen is a forgery, that person may never lose completely the lingering negative impression the video has left with them.

It is also the case that not only may fake videos be passed off as real, but real information can be passed off as fake. This is the so-called "liar's dividend" in which people with a propensity to deceive are given the benefit of an environment in which it is increasingly difficult for the public to determine what is true.

To give our members and the public a sense of the quality of deepfakes today, I want to share a few short examples, and even these are not the state of the art. The first comes from Bloomberg Businessweek and demonstrates an AI-powered cloned voice of one of the journalists. So let's watch the clip.

[Video shown]

The Chairman. All right. It is bad enough that was a fake, but he is deceiving his mother and telling her that he has got a virus. That seems just downright cruel.

The second clip comes from Quartz and demonstrates a puppet master type of deepfake video.

[Video shown.]

The Chairman. As you can see, these people are able to co-op the head movements of their targets. If married with convincing audio, you can turn a world leader into a ventriloquist's dummy.

Next, a brief CNN clip highlighting new research from Professor Hany Farid, an acclaimed expert on deepfakes from UC Berkeley, and featuring an example of a so-called

"face swap" video in which Senator Elizabeth Warren's face is seamlessly transponded on the body of SNL actress Kate McKinnon.

[Video shown]

The Chairman. So the only problem with that video is Kate McKinnon actually looks a lot like Elizabeth Warren, but the one on the left was actually -- both were Kate McKinnon, one just had Elizabeth Warren's face swapped onto her, but it shows you just how convincing that kind of technology can be.

These algorithms can also learn from pictures of real faces to make completely artificial portraits of persons who do not exist at all.

[Video shown.]

The Chairman. Can anyone here pick out which of these faces are real and which are fake? And, of course, as you may all have guessed, all four are fake. All four of those faces are synthetically created, none of those people are real.

Think ahead to 2020 and beyond. One does not need any great imagination to envision even more nightmarish scenarios that would leave the government, the media, and the public struggling to discern what is real and what is fake. A state-backed actor creates a deepfake video of a political candidate accepting a bribe with the goal of influencing an election; or an individual hacker claims to have stolen audio of a private conversation between two world leaders when, in fact, no conversation took place; or a troll farm uses text-generating algorithms to write false or sensational news stories at scale, flooding social media platform and overwhelming journalists' ability to verify and users' ability to trust what they are seeing or reading.

What enables deepfakes and other modes of disinformation to become truly pernicious is the ubiquity of social media and the velocity at which false information can spread. We got a preview of what that might look like recently when a doctored video

of Speaker Nancy Pelosi went viral on Facebook, receiving millions of views in the span of 48 hours.

That video was not an AI-assisted deepfake, but rather a crude manual manipulation that some have called a cheapfake. Nonetheless, the video's virality on social media demonstrates the scale of the challenge we face and the responsibilities that social media companies must confront.

Already the companies have taken different approaches, with YouTube deleting the altered video of Speaker Pelosi, while Facebook labeled it as false and throttled back the speed it spread once it was deemed fake by independent fact checkers.

Now is the time for social media companies to put in place policies to protect users from this kind of misinformation, not in 2021, after viral deepfakes have polluted the 2020 elections, by then it will be too late.

And so, in keep with a series of open hearings that have examined different strategic challenges to our national security and our democratic institutions, the committee is devoting this hearing to deepfakes and synthetic media.

We need to soberly understand the implications of deepfakes, the underlying AI technologies, and the internet platforms that give them reach before we consider appropriate steps to mitigate the potential harms.

We have a distinguished panel of experts and practitioners to help us understand and contextualize the potential threat of deepfakes. But before turning to them, I would like to recognize Ranking Member Nunes for any opening statement he would like to give.

[The statement of The Chairman follows:]

******** COMMITTEE INSERT ********

Mr. Nunes. Thank you, Mr. Chairman. I join you in your concern about deepfakes, and I want to add to that, fake news, fake dossiers, and everything else that we have in politics.

I do think that, in all seriousness, though, this is real. If you get online you can see pictures of yourself, Mr. Chairman, on there. They are quite entertaining, some of them. I decided not to play --

The Chairman. Maybe they are entertaining for you.

Mr. Nunes. I decided not to play them today on the screen.

But with all seriousness, I appreciate the panelists being here and look forward to your testimony. I yield back.

[The statement of Mr. Nunes follows:]

******** COMMITTEE INSERT ********

The Chairman. I thank the ranking member. Without objection, these opening statements will be made part of the record. I would like to welcome today's panel.

First, Jack Clark, who is the policy director of OpenAI, a research and technology company based in San Francisco, and a member of the Center for New American Security's Task Force on Artificial Intelligence and National Security.

Next, David Doermann, a professor and director or artificial intelligence, the Artificial Intelligence Institute at the University of Buffalo. Until last year he was the program manager of DARPA's media forensics program.

Danielle Citron is a professor of law at the University of Maryland Francis King Carey School of Law, and she has coauthored several notable articles about the potential impacts of deepfakes on national security and democracy.

And finally, Mr. Clint Watts, who is a distinguished research fellow at the Foreign Policy Research Institute, a senior fellow at GMF's Alliance for Securing Democracy, and his recent scholarship has addressed social media influence operations.

Welcome to all of you.

And why don't we start with you, Mr. Clark?

STATEMENTS OF JACK CLARK, POLICY DIRECTOR, OPENAI; DR. DAVID DOERMANN, PROFESSOR, SUNY EMPIRE INNOVATION, AND DIRECTOR, ARTIFICIAL INTELLIGENCE INSTITUTE, UNIVERSITY AT BUFFALO; DANIELLE CITRON, PROFESSOR OF LAW, UNIVERSITY OF MARYLAND FRANCIS KING CAREY SCHOOL OF LAW; AND CLINT WATTS, DISTINGUISHED RESEARCH FELLOW, FOREIGN POLICY RESEARCH INSTITUTE, AND SENIOR FELLOW, ALLIANCE FOR SECURING DEMOCRACY, GERMAN MARSHALL FUND

STATEMENT OF JACK CLARK

Mr. Clark. Chairman Schiff, Ranking Member Nunes, and committee members, thank you for the invitation to testify about the national security threats posed by the intersection of AI, fake content, and deepfakes.

So what are we talking about when we discuss this subject? Fundamentally, we are talking about digital technologies that make it easier for people to create synthetic media, and that can be video, images, audio, or text.

Now, people have been manipulating media for a very long time, as you well know, but things have changed recently. And I think there are two fundamental reasons for why we are here.

One is that the continued advancements of computing capabilities, that is the physical hardware we use to run software on, that has got significantly cheaper and more powerful. And at the same time software has become increasingly accessible and capable, and some of this software is starting to incorporate AI, which makes it dramatically easier for us to manipulate media, and it allows for a step change in functionality for things like video editing or audio editing, which was previously very

difficult.

Now, the forces driving cheaper computing and easier-to-use software are fundamental to the economy and many of the innovations that we have had in the last few years. So when we think about AI, one of the confounding factors here is that similar AI technologies used in the production of synthetic media or deepfakes are also likely to be used in valuable scientific research. They are used by scientists to allow people with hearing issues to understand what other people are saying to them, or they are used in molecular assay and other things that which may revolutionize identified medicine.

Now, at the same time this techniques can be used for purposes that justifiably cause unease, like being able to synthesize the sound of someone else's voice, impersonate them on video, and write text in the style they use on line. We have also seen researchers develop techniques that combine these things, allowing them to create a virtual person who can say things that they haven't said and appear to do things that they haven't necessarily done.

I am sure that members of the committee are familiar with their run-ins with the media and know just how awkward it can be to have words put in your mouth that you didn't say. So deepfakes take this problem and potentially accelerate it.

So how might we approach this challenge? I actually think there are several interventions that we can make, and this will improve the state of things.

One is institutional interventions. It may be possible for large-scale technology platforms to try and develop and share tools for the detection of malicious synthetic media at both the individual account level and the platform level, and we could imagine these companies working together privately, as they do today with cybersecurity, where they exchange threat intelligence with each other and with other actors to develop a

shared understanding of what this looks like.

We can also increase funding. So, as mentioned, Dr. David Doermann previously led the DARPA program here. We have existing initiatives that are looking at the detection of these technologies, and I think that it would be judicious to consider expanding that funding further so that we can develop better insights here.

I think we can measure this. And what I am mean by measurement is that it is great that we are here now ahead of 2020, but these technologies have been in open development for several years now, and it is possible for us to read research papers, read code, talk to people, and we could have created quantitative metrics for the advance of this technology for several years.

And I strongly believe that government should be in the business of measuring and assessing these threats by looking directly at the scientific literature and developing a base of knowledge from which to work out next steps. Being forewarned is forearmed here, and we can do that.

I think we also need to do work at the level of norms. So at OpenAI we have been thinking about different ways to release or talk about the technology that we develop. I think that it is challenging because science runs on openness, and we need to preserve that so that science continues to move forward, but we do need to consider different ways of releasing technology or talking to people about the technology that we are creating ahead of us releasing it.

Finally, I think we need comprehensive AI education. None of this works if people don't know what they don't know. And so we need to give people the tools to let them understand that this technology has arrived, and though we may make a variety of interventions to deal with the situation, they need to know that it exists.

So as I hope this testimony has made clear, I don't think AI is the cause of this. I

think AI is an accelerant to an issue that has been with is for some time. And we do need to take steps here to deal with this problem because the pace of this is challenging.

Thank you very much.

[The statement of Mr. Clark follows:]

******** INSERT 1-1 ********

The Chairman. Thank you.

Mr. Doermann.

STATEMENT OF DAVID DOERMANN

Mr. Doermann. Thank you. Chairman Schiff, Ranking Member Nunes, distinguished members of the committee, thank you for the opportunity to be here this morning to discuss the challenges of countering media manipulation at scale.

For more than five centuries authors have used variations of the phrase "seeing is believing." But in just the past half-decade, we have come to realize that that is no longer always true.

In late 2013, I was given the opportunity to join DARPA as a program manager and was able to address a variety of challenges facing our military and our intelligence communities. Although I am no longer a representative of DARPA, I did start the media forensics program -- MediFor -- and it was created to address the many technical aspects about the problems that we are talking about today.

The general problem of MediFor is addressing our limited ability to analyze, detect, and address manipulated media that, at the time, was being used with increased frequently by our adversaries.

It is clear that our manual processes, despite being carried out by exceptionally competent analysts and personnel in the government, at the time could not deal with the problem at scale that the manipulated content was being created and proliferated.

In typical DARPA fashion, the government got ahead of this problem, knowing that it was a marathon, not a sprint, and that the program was designed to address both current and evolving manipulation capabilities, not with a single point solution, but with a

comprehensive approach.

What was unexpected, however, was the speed at which this manipulation technology would evolve. In just the past 5 years we have gone from a new technology that could produce novel results at the time, but nowhere near what could be done manually with basic desktop editing software, to open source software, such as deepfakes, that can make the manual effort completely out of the equation.

Now, there is nothing fundamentally wrong with or evil about the underlying technology that gives rise to the concerns that we are testifying about today. Like basic image and video desktop editor, deepfakes is only a tool, and there are a lot more positive applications of generative networks than there are negative ones.

As of today, there are point solutions that can identify deepfakes reliably, but it is only because the focus of those developing the deepfakes-like technology have been on visual deception, not on covering up trace evidence. If history is any indicator, it is only a matter of time before the current detection capabilities will be rendered less effective, in part because some of the same mechanisms that are used to create this content are also used to cover them up.

I want to make it clear, however, that combating synthetic and manipulated media at scale is not just a technical challenge, it is a social one as well, as I am sure other witnesses will be testifying this morning. And there is no easy solution, and it is likely to get much worse before it gets much better. Yet we have to continue to do what we can.

We need to get the tools and the processes in the hands of individuals rather than relying completely on the government or on social media platforms to police content. If individuals can perform a sniff test and the media smells of misuse, they should have ways to verify it or prove it or easily report it.

The same tools should be available to the press, to social media sites, to anyone

who shares and uses this content, because the truth of the matter is the people that share this stuff are part of the problem, even though they don't know it.

We need to continue to work towards being able to apply automated detection and filtering at scale. It is not sufficient to only analyze questioned content after the fact, we need to be able to apply detection at the front end of the distribution pipeline.

And even if we don't take down or prevent manipulated media from appearing, we should provide appropriate warning labels that suggest that this is not real or not authentic or not what it is purported to be. And that is independent of whether this is done and the decisions are made by humans, machines, or a combination.

And we need to continue to put pressure on social media to realize that the way that their platforms are being misused is unacceptable. They must do all they can to address today's issues and not allow things to get worse.

Let there be no question that this is a race. The better manipulators get, the better detectors need to be. And there is certain orders of magnitude more manipulators than detectors.

It is also a race that may never end. It may never be won. But it is one where we must close the gap and continue to make it less attractive financially, socially, politically, to propagate false information. Like spam and malware, it is easy and it is always a problem, and it may be the case that we can level the playing field.

When the MediFor program was conceived at DARPA, one thing that kept me up at night was the concern that someday our adversaries would be able to create entire events with minimal effort. These events might include images of scenes from different angles, video content that appears from different devices, and text that is delivered through various media, providing an overwhelming amount of evidence that an event has occurred, and this could lead to social unrest or retaliation before it gets countered. If

the past 5 years are any indication, that someday is not very far in the future.

Thank you.

[The statement of Mr. Doermann follows:]

******** INSERT 1-2 ********

The Chairman. Thank you.

Professor Citron.

STATEMENT OF DANIELLE CITRON

Ms. Citron. Thank you, Chairman Schiff, Ranking Member Nunes, and the committee, for having me her today to talk about the phenomenon of deepfakes, the risks that they pose and what law canon should do about it. So I am a professor of law at the University Maryland School of Law.

And there are a few phenomena that come together that make deepfakes particularly troubling when they are provocative and destructive. The first is that we know that as human beings, the video and audio is so visceral, we tend to believe what our eyes and ears are telling us. And we also tend to believe and tend to share information that confirms our biases. And it is particularly true when that information is novel and negative. So the more salacious, we are more willing to pass it on.

And we are seeing deepfakes, or will see them, in social networks that are ad-driven. So the entire enterprise is to have us click and share. So when we bring all of these things together, the provocative deepfake, the salacious will be spread virally.

So let me describe, there are so many harms that my coauthor Bobby Chesney and I have written about, but I am going to focus on some of the more concrete ones, and what law canon should do about it.

So there are concrete harms in the here and the now, and especially for individuals. So Rana Ayyub is an investigative journalist in India who writes about government corruption and the persecution of religious minorities. And she is long used to getting death threats and rape threats. For her, it is sort of par for the course.

But she wrote a provocative piece in April 2018, and what followed was posters circulated over the internet deepfake sex videos of Rana. So her face was morphed into pornography.

And that first day it goes viral, it is on every social media site, WhatsApp, it is, as she explained to me, millions of phones in India. And the next day -- so paired with the deepfake sex video of Rana, was rape threats, her home address, and the suggestion that she was available for sex.

Now, the fallout was significant. She had to basically go offline. She couldn't work. Her sense of safety and security was shaken. It upended her life. And she had to withdraw from online platforms for several months. So the economic and the social and the psychological harm is profound. And it is true that, in my work on cyber stalking, the phenomenon is going to be increasingly felt by women and minorities and for people from marginalized communities.

Now, of course, it is not just individuals. We can imagine the deepfake about sort of the night before an IPO, timed just right, with the CEO saying something that he never said or did, basically admitting to the company was insolvent. And so the deepfake, if the night before the IPO could upend the IPO, the market will respond far faster than we can debunk it.

So the question is -- and we can imagine all sorts of -- and Mr. Watts and I have talked about -- I am going to let him take some of the national security concerns, like elections, the tipping of an election, upending public safety. But the next question is, what do we do about it?

And I feel like our panel is going to be in heated agreement that there is no silver bullet, that we need a combination of law, markets, and really societal resilience to get through this.

But law has a modest role to play. There are civil claims that victims, targeted individuals can bring. They can sue for defamation, intentional infliction of emotional distress, false light, a privacy tort. But the hardest thing is that it is incredibly expensive to sue.

And criminal law offers too few levers for us to push. At the State level there are a handful of criminal defamation laws and impersonation laws, and at the Federal level there is an impersonation of a government official statute, but it is really inapt for the sets of problems that we face today.

And so Professor Mary Anne Franks and I are amidst writing a model statute that we might deploy, one that is narrowly tailored, that would address harmful false impersonations, that would capture some of the harm here. But, of course, there are practical hurdles for any legal solution. You have to be able to find the defendant to so prosecute them and you have got to have jurisdiction over them.

And the platforms, the intermediaries, our digital gatekeepers are immune from liability. So we can't use a legal incentive of liability to get them on the case.

So I see my time is running out. I look forward to your questions. And thank you.

[The statement of Ms. Citron follows:]

******** INSERT 1-3 ********

The Chairman. Thank you very much.

Mr. Watts.

STATEMENT OF CLINT WATTS

Mr. Watts. Chairman Schiff, Ranking Member Nunes, members of the committee, thanks for having me here today.

All advanced nations recognize the power of artificial intelligence to revolutionize economies and empower militaries, but those countries with the most advanced AI capabilities and unlimited access to large data troves will gain enormous advantages and information warfare.

AI provides purveyors of disinformation the ability to rapidly recon American social media audiences, to identify psychological vulnerabilities, and to create modified content and digital forgeries advancing false narratives against Americans and American interests.

Historically, each advancement in media, from text, to speech, to video, to virtual reality, more deeply engages information consumers, enriching the context of experiences and shaping a user's reality. The falsification of audio and video allows manipulators to dupe audience members in highly convincing ways, provoking emotional responses that can lead to widespread mistrust and, at times, physical mobilizations. False video and audio, once consumed and believed, can be extremely difficult to refute and counter.

Moving forward, I would estimate Russia, as an enduring purveyor of disinformation, is and will continue to pursue the acquisition of synthetic media capabilities and employ the outputs against its adversaries around the world. I suspect

they will be joined and outpaced potentially by China. China's artificial intelligence capabilities rival the U.S., are powered by enormous data troves, to include vast amounts of information stolen from the U.S., and the country has already shown a propensity to employ synthetic media in television broadcast journalism.

These two countries, along with other authoritarian adversaries and their proxies, will likely use deepfakes as part of disinformation campaigns seeking to discredit domestic dissidents and foreign detractors, incite fear and promote conflict inside Western-style democracies, and, three, distort the reality of American audiences and the audiences of America's allies.

Deepfake proliferation presents two clear dangers. Over the long term, deliberate development of false synthetic media will target U.S. officials, institutions, and democratic processes with an enduring goal of subverting democracy and demoralizing the American constituency.

In the near and short term, circulation of deepfakes may incite physical mobilizations under false pretenses, initiating public safety crises and sparking the outbreak of violence. The recent spate of false conspiracies proliferating via WhatsApp in India offer a relevant example of how bogus messages in media can fuel violence. The spread of deepfake capabilities will only increase and the frequency and intensity of these violence outbreaks will continue.

U.S. diplomats and military personnel deployed overseas will be prime targets for deepfake disinformation conspiracies planted by adversaries. U.S. interests in the developing world, where consumption has jumped from analog in-person conversations to social media sharing lacking any form of verification filter, will likely be threatened by bogus synthetic media campaigns.

Three examples would be mobilization at the U.S. Embassy in Cairo, the consulate

in Benghazi, and rumors of protests at Incirlik Air Base, had they been accompanied with fake audio or video content, could have been far more damaging in terms of that.

I would also point to a story just out hours ago from the Associated Press which shows the use of a synthetic picture for what appears to be espionage purposes on LinkedIn, essentially a honey potting attack.

Recent public discussions of deepfake employment heavily focus on foreign adversaries, but the greatest threat of inauthentic content proliferation may come not from abroad but from home, and not from nation-states but from the private sector.

Thus far, I have focused on authoritarian nation-states. I brought a chart here today. But a range of advanced persistent manipulators will use their vast resources to develop and acquire deepfakes as needed in pursuit of their goals.

Recent examples of disinformation and misinformation suggest it could be oligarchs, corporations, political action groups, public relation firms, and activists with significant financial support that will seek out these media capabilities and amplify deepfakes in international or domestic contexts.

The net effect will be the same: degradation of democratic institutions and elected officials, lowered faith in electoral processes, weakened trust in social media platforms, and potentially sporadic violence by individuals and groups mobilized under false pretenses.

I have several recommendations, but I will only hit a couple here in the oral remarks.

First, Congress should implement legislation prohibiting U.S. officials, elected representatives, and agencies from creating and distributing false and manipulated content. The U.S. Government must always be the purveyor of facts and truth to its constituents, assuring the effective administration of democracy via productive policy

debate from a shared basis of reality.

Second, policymakers should work jointly with social media companies to develop standards for content accountability.

Third, the U.S. Government should partner with the private sector to implement digital verification signatures designating the date, time, and physical origination of content.

Fourth, social media companies should enhance their labeling of synthetic content across platforms and work as an industry to codify how and when manipulated or fake content should be appropriately marked. Not all synthetic media is nefarious in nature, but information consumers should be able to determine the source of the information and whether it is an authentic depiction of people and events.

Fifth, and what I think is the most pressing right now, is the U.S. Government from a national security perspective should maintain intelligence on adversaries capable of deploying deepfake content or the proxies they employ to conduct such disinformation. The Departments of Defense and State should immediately develop response plans for deepfake smear campaigns and deepfake-inspired violent mobilizations overseas in an attempt to mitigate harm to U.S. personnel and interests.

And the last, I echo my fellow panelists, is public awareness of deepfakes and its signatures will greatly assist in tamping down attempts to subvert U.S. democracy and incite violence. I would like to see us help the public make better decisions about the content that they are consuming and how to judge that content.

Thank you.

[The statement of Mr. Watts follows:]

******** INSERT 1-4 ********

The Chairman. Thank you all.

We will now proceed with questions. I recognize myself for 5 minutes.

Two questions, one for Professor Citron and one for Mr. Watts.

Professor, how broad is the immunity that the social media platforms enjoy? And is it time to do away with that immunity so the platforms are required to maintain a certain standard of care?

It seems to me not very practical to think about bringing people to justice who are halfway around the world, or the difficulties of attribution, or the fact that, given the cheap cost of this technology now, just how much people can employ it.

Is it time to take that step?

Was it appropriate for one social media company to leave up the Pelosi video, even labeling it in a certain way?

And, Mr. Watts, what is a proportionate response should the Russians start to dump deepfakes, release a deepfake of Joe Biden to try to diminish his candidacy? What should the U.S. response be? Should it be a cyber response, not a tit for tat, in the sense of doing a deepfake of Putin, but rather some cyber reaction, or are sanctions a better response? How do we deter this kind of foreign meddling, realizing that that is only going to be one part of the problem?

Professor.

Ms. Citron. So I am going to start with how broad the immunity is, and then that it is time for us to amend Section 230 of the Decency Act.

So under a law passed in 1996, the Communications Decency Act, it largely was an anti-porn provision. I mean, if we can imagine the internet without porn. That was the objective of the Communications Decency Act.

And most of that law is struck down, but what remains is a provision, it is called

Good Samaritan Blocking and Filtering of Offensive Content. And it has been interpreted really broadly to say that if you under-filter content, if you don't engage in any self-monitoring at all, even if you encourage abuse, that you are immune from liability for user-generated content.

So that means that revenge porn operators can gleefully say that they are immune from liability while encouraging people to post their ex's nude photos. And they are right, they are immune from liability, because they are not generating, co-creating the content.

So the question is, here we are 25 years later, the internet, we have got dominate players, it is not -- the internet is not in its infancy, and is it time to reassess it?

And I think the answer is yes, that we should condition the immunity, it shouldn't be a free pass, and it should be conditioned on reasonable content moderation practices. Ben Wittes and I have written sort of a sample statute that you could adopt if you so chose that would condition the immunity on reasonable content practices.

Then the question of course is, well, in any given case are platforms making the right choices? And under an approach that would have us look to the reasonableness of content practices, we would look at the platform's total, its approach generally speaking to content moderation, not any given decision with content.

So let's take the Pelosi video. I think the answer is it should have been taken down. We should have a default rule, the platform should have a default rule, that if we are going to have impersonations or manipulation that do not reflect what we have done or said, then platforms should, once they figure it out, take it down.

The technology is such that we can't detect it yet, we can't automatically filter and block. But once we have figured it out, we already are in a place where the public has deep distrust of the institutions at the heart of our democracy and you have an audience

primed to believe things like manipulated video of lawmakers.

And I would hate to see the deepfake where a prominent lawmaker purported to show seen taking a bribe that you never took. And I hope that platforms come to see themselves, if we can't require them to have legal liability, that they come to see themselves as the purveyors responsibly of facilitating discourse online and their importance to democracy.

The Chairman. Thank you.

Mr. Watts.

Mr. Watts. Yeah, I would like to start off with just a basic principle of information warfare. And so R.H. Knapp, who was a professor who essentially wartime rumors, his quote was: "Once rumors are current, they have a way of carrying the public with them. The more a rumor is told, the greater its plausibility." And he wrote that in 1944. I think that is still the essential thing. It comes down, who is there first and who is there the most? And that is the danger of social media computational propaganda with this AI.

In terms of how we deal with this, there are several parts. One is we have to have a plan, and it is a multipart plan. The other part is we have to respond quickly. This has not been the tradition of our government. For example, in Iraq, when there would be fake al-Qa'ida propaganda put out to try and inspire people to show up places, we had rapid response teams that would show up with video, with audio, that would shoot footage from there to show this is not true, this has been disproven.

That is a great example about, if this starts to get leaked out, what is our plan right now? The U.S. Government, for any government official, government agency, should immediately offer a counter based on fact in terms of what is actually going on.

This happened in the summer of 2016 at Incirlik Air Base. There was Russian

state-sponsored propaganda put out about a potential coup, maybe the base was surrounded, maybe there is a protest. We should be able to turn on the cameras at that air base immediate and say this is not happening. The faster we do that, the less chance people see it first, the less chance that people see it often and believe it.

The second part is I think it comes down to the political parties, Republican and Democrat. If they have these smears coming through, they should be able to rapidly refute that and put out a basis of truth, this candidate or candidates were not there. But that means partnership also with the social media companies in terms of this.

I actually would not go as far as saying every piece of synthetic video that gets loaded up on a social media platform needs to come down.

And I am glad you brought up former Vice President Biden. One of the classic articles about former Vice President Biden comes from The Onion, and it was that he was waxing his Camaro in the driveway of the White House. It was a comedy bit and it had manipulated photos, manipulated content on it.

If we went to that extreme, we would have a country where everything that has ever been changed or modified for any reason would have to be policed, and we would be asking a private sector company to police that.

So I would instead offer a different system, which is triage, which is social media companies, how do they accurately label content as authentic or not. The danger with social media is the source is not necessarily there. We saw that in 2016, we see that today.

So they should be able to refer back to whatever the base source is quickly. How do you do that? They should be able to triage.

The three areas which I would suggest that they immediately triage in is if they see something spiking in terms of virality they should immediately put that into a queue

and have it done for human review. Link to fact-checkers, downrate it, not let it go into news feeds, and help the mainstream media also understand what is manipulated content. That is the jump that we are most concerned about.

The other part is outbreaks or potential outbreaks of violence and public safety. And then anything related to elected officials or public institutions, should immediately be flagged and pulled down and checked and then a context be given to it.

I see it as the public needs to be given a context that we are not really suppressing all freedom of speech, all development of this context, because there are legitimate reasons that we might want to use synthetic media for entertainment, comedy, all sorts of visualizations that are out there.

The Chairman. Okay. We will go to Mr. Nunes, but at some point I would love to follow up see what a proportionate response would be to a foreign adversary that deploys this kind of --

Mr. Watts. I can actually, if you give me 20 seconds, I can tell you what it would be, which is refuting, number one. Number two, I think offensive cyber is in place, and I like what the NSA has actually done in 2018. And then, number three, more aggressive responses in terms of sanctions. I like sanctions around troll forms and these cut-outs where this content comes from.

The Chairman. Thank you.

Mr. Nunes.

Mr. Nunes. Thank you, Mr. Chairman.

So how do you put in filters to these tech oligarch companies -- there are only a few of them, you know who they are -- that doesn't -- it is not developed by partisan left wing, like it is now, where most of the time it is conservatives who get banned and not Democrats? Like the Pelosi video was taken down, that is fine, I don't have a problem

with that. But I can tell you there are videos up of Republicans that go on and on and on. So it is all in who is building the filter, right?

Ms. Citron. Are you asking somebody --

Mr. Nunes. Well, you were the one that was talking about filters, so --

Ms. Citron. Yeah. No. And what I was suggesting is that it would be impossible to ex ante filter deepfake content. We really can't detect it as far as the state-of-the-art goes now, nor do I think in the arms race will we be able to really filter it.

And what I was saying is that for something like a video, where it is clearly doctored and an impersonation, not satire, not parody. There are wonderful uses for deepfakes that are art, historical, sort of rejuvenating for people to create them about themselves. So I am not suggesting all deepfakes, but rather --

Mr. Nunes. No, and I am not, but I think I mostly agree with you, other than I just don't know how you -- I think the challenge here is how do you implement it?

Ms. Citron. Right.

Mr. Nunes. Which is what you have been studying, it sounds like.

Ms. Citron. And these are really hard problems of content moderation. I have worked with companies for about 10 years now, and in particular on the issue of nonconsensual pornography, threats, and stalking, and it is such a contextual question.

And so you can't proactively filter. But when it is reported, the question is -- and when we see videos going viral, there is a way in which we -- companies should react, and react responsibly.

And absolutely should be bipartisan, there shouldn't be ideology that drives the question, but rather is this a misrepresentation in a defamatory way, that we would say it is a falsehood that is harmful to reputation, that is an impersonation? Then we should take it down. So I think that is the default I am imagining for social media companies,

but it would be ex post.

Mr. Nunes. But it is a challenge. I mean, you talked about the 1996 law that needs to be changed. And I think it has to be one way or another, right? Either they have to be truly an open public square, which then it is very difficult to filter because then whoever is developing the filter puts their own bias into the filter.

Ms. Citron. But actually 1996, that bill, it did not imagine an open public square where private companies couldn't filter. The opposite. It was designed to encourage self-monitoring and to provide an immunity in exchange for Good Samaritan filtering and blocking of offensive content.

So the entire premise of Section 230 is to encourage and so provide an immunity so that there was filtering and blocking, because Congress knew it would be too hard for Congress or the FTC to get ahead of all this themselves. And that was in 1996. Imagine now the scale that we face.

I think we should preserve the immunity, but condition it on reasonable content moderation practices so that there are some sites that literally traffic in abuse, that encourage illegality, and they should not enjoy immunity from liability.

Mr. Nunes. Right. But then we are back to where we started. I mean, this is the challenge, right? So how do we draft legislation that would --

Ms. Citron. Yep.

Mr. Nunes. -- that would enable that --

Ms. Citron. Happy to tell you how to do it.

So Section 230(c)(1) now says: No speaker or publisher -- or no online service shall be treated as the speaker or publisher, essentially, of someone else's content. What we can do is change Section 230(c)(1) to say that no online service that engages in reasonable content moderation practices shall be treated as the speaker or publisher of

somebody's content. So we can change Section 230 with some imagination.

Mr. Nunes. Then it depends on what the definition of reasonableness is.

Ms. Citron. And that is what law does really well. So every time I hear a lawyer say we can't figure out what is reasonable, it is called tort law. Negligence is built on the foundation of reasonableness. You know, so often law moves in a pendulum. We often start with no liability because we really want to protect businesses, and we should, and we experiment and we realize there is a lot of harm. And then we also overreact and impose strict liability. And then we get somewhere in the middle. That is where negligence lives, reasonable practices.

And we have industries. Content moderation has been going on for the past 10 years, and I have been advising Twitter and Facebook of all that time. There is meaningful reasonable practices that are emerging and have emerged in the last 10 years.

So we have a guide. It is not as if this is a new issue in 2019. So we can come up with reasonable practices.

Mr. Nunes. Thank you. I yield back, Mr. Chairman.

The Chairman. Mr. Himes.

Mr. Himes. Thank you, Mr. Chairman.

Dr. Doermann, I want to get a quick sense from you of what the status quo is with respect to our ability to detect and where that race is.

But before I do that, I just want to highlight something that I think is of actually very immediate and intense interest to the Intelligence Community.

Mr. Watts, you said something which is -- if something is happening on a base somewhere, we can just turn on the cameras. I am not sure that is right, right? Because if you can create a deepfake, there is no reason why you can't create a deepfake

from that camera to the screen, right?

So, in other words, the point I am trying to make is our Intelligence Community obviously relies on things like full motion video and photographs and that sort of thing. One of the threats here is not just the threats that we might be made to look silly on YouTube, but that our Intelligence Community, using its own assets, might not be able to tell fact from fiction. Is that correct? When you say let's just turn on the cameras, I am not sure that is enough.

Mr. Watts. Well, I think it needs to be -- one of my other recommendations was digital verification, which these folks will know better because they are more technically sound on this than I am. But digital verification for date, time, and location of actual content to include realtime content. There are already some blockchain registry solutions that are being developed. That is essential.

Part of that would be then, if you, as the U.S. Government, turn on your cameras, it can be verified by news agencies, reporters, we could have it on C-SPAN. We could use it in a lot of different ways. But we have to make sure that we have the ability to verify that our content is real, so if that sort of an impersonation is done, we can do it quickly and people will know which one to sort through.

I will defer to them in terms of the technical, but some of this is already being developed. It is not quite there yet. But I would want that accompanied with it.

Mr. Himes. Great. Thank you. That actually leads into my question for Dr. Doermann.

Dr. Doermann, I understand there is no silver bullet here, this is going to be a cat-and-mouse game. Take a minute or 2 and just tell where are we in that cat-or-mouse game? Should we expect to have undetectable deepfakes out there within a year, 2 years, 5 years, 10 years? Where are we today and how imminent a

challenge is this?

Mr. Doermann. I think there is the risk of having undetectable content that gets modulated on -- shared online. Right now things like compression, if you have a very low resolution version of a video, the attribution can be destroyed. The camera fingerprint where this content came from can be destroyed. A lot of the trace evidence could be destroyed with very simple types of manipulation on top of the deepfake process, or any type of manipulation.

The challenge that we have, though, is that we do have point solutions for a lot of these components, and bringing them together in a useful way, and as I said, getting them in the hands of everyone throughout the pipeline.

Imagine if Facebook or YouTube or any of these other companies could have this up front, and when the human reviewers -- Facebook, I think, just reported they have hired 30,000 people to review content -- could have this ahead of time, rather than saying, okay, I have a questioned video or a questioned piece of audio or something that I need to review, now let me go and run this algorithm on it or this set of algorithms. Do that up front so they have a report associated with a particular image or video. And then if there are questions, to put that warning up there.

I think the public doesn't know enough about what is possible to demand that if somebody knows something. The truth of the matter is when this stuff gets shared -- it gets created once, and when it gets shared, it gets shared across different platforms. It gets shared by different people with different media.

But the truth of the matter is, that signature for that particular piece of video, that piece of audio, is there. And so there are tools that the social media companies could use to link those together and make a decision and then share it with everyone, the same way as we do with malware, for example, cyber issues.

We have gotten to the point where it is now, you know, we are protecting our front door, and we need to protect our front door from images and video as well.

Mr. Himes. Thank you, Doctor.

Professor Citron, I don't have time to cover this topic, but I just want to express myself on this. The theme of this hearing is how scary deepfakes are, but I got to tell you, one of the more scary things I have heard this morning is your statement that the Pelosi video should been taken down, there should be this.

I don't have a lot of time, sadly, there won't be a moment, I think, for you to answer, but I do want to have this conversation, because as awful as I think we all thought that Pelosi video was, there has got to be a difference if the Russians put that up, which is one thing, versus if Mad magazine does that as a satire.

As you know better than me, we don't have a lot of protections as public figures with respect to defamation. And some of the language you have used here today makes me worry about First Amendment equities, free expression, centuries-long tradition of satirizing people like us who richly deserve being satirized.

So anyway, I am expounding here. I simply just wanted to put that on the record and hope that we have an opportunity this morning to hear more about where that boundary lies and how we can protect that long tradition of freedom of expression.

With that, I will yield back.

The Chairman. Dr. Wenstrup.

Dr. Wenstrup. Thank you, Mr. Chairman.

Thank you all for being here.

I think, boy, we have come a long way. I remember Chevy Chase playing Gerald Ford on "Saturday Night Live" and they didn't even try to pretend to look like Gerald Ford. And then we see "Forrest Gump," which was, you know, a wonderful movie, right, it was

entertainment. And I remember sitting there thinking, how did they do that?

You know, the problem we have -- I have always said that out of everything bad there is a chance to do something good, but out of everything good there is obviously a chance for people to do something bad. And I think that we see that. And the way it sounds with where we are headed, it is like we are all living in "The Truman Show" or something like that, and we have got to be careful about that.

But I think about, in that vein of out of something good something bad can happen, I am sure the Wright brothers when they learned to fly didn't think: And maybe we can fly this into a building some day and kill people. Right? But that is what happens in this world, unfortunately. But as a result of that, 9/11, for example, it takes a lot longer to get on a plane, and for good reason.

And I think that where we need to be headed might be, and I want your opinions on it, obviously, we have got to slow this down, you know, before something just hits it. And I think you are talking about this, sort of the triage idea.

And maybe we label. Maybe, unfortunately, we have to tell people before they see something, this is satire, it is not real, and you have to in some way verify, which is kind of pathetic, but at the same time that may be what we have to do. Slow it down. Triage it. This is not verified. This is satire.

And maybe on a global scale, when it comes to punitive measures, the people that are doing things nefarious, maybe we have to have international extradition laws, because when something comes from some other country, maybe even a friendly country, that defames and hurts someone here, maybe we both agree amongst those nations that we will extradite those people and they can be punished in your country for what they did to one of your citizens.

So I would love your opinion on those, the triage, labeling, and extradition.

Whoever wants to take it first.

Mr. Doermann. Yeah. I think that is absolutely right. I mean, one of the reasons that these types of manipulated images and video gain traction is because it is almost instantaneous that they can be shared. They can be shared around the world. They can be shared across platforms. You can see something on one platform and there is a button there to post it to another.

There is an old adage that says that a lie can go halfway around the world before the truth can get its shoes on, and that is true.

Dr. Wenstrup. Triage.

Mr. Doermann. Personally, I don't see any reason why. Broadcast news does it with live types of -- you know, they have a delay, a 7-second delay or a 15-second delay. There is no reason why things have to be instantaneous. Our social media should instill these types of things, to delay, so that they can get these types of things online. They can decide whether they should label it. We still need to put the pressure on for those types of things.

There is a seriousness issue. There is from satire all the way up through child pornography. We have done it for child pornography, we have done it for human trafficking, they are serious about those things. This is another area that is a little bit more in the middle, but I think they can make the same effort in these areas to do that type of triage.

Dr. Wenstrup. Yeah. If you say, what you are about to see is satire and has been modified.

Go ahead.

Mr. Clark. So I think one thing worth stressing is we will continue to be surprised by technological progress in this domain, because the lure of a lot of this stuff is all of

these people think they are the Wright brothers and they feel that and they are all busily creating stuff. Figuring out the second order of effects of what they build is difficult.

So I think that we do need to build infrastructure so that you have some third party measuring the progression of these technologies so you can anticipate the other things in expectation.

Dr. Wenstrup. Ms. Citron.

Ms. Citron. And the labeling, I think, is incredibly important, and there are times in which that is the most -- that is the perfect, rather than second best, where we should err on the side of inclusion and label it as synthetic and be so required to label it.

And it is true that there are some instances, though, where we say where labeling is just not good enough, that it is defamatory, that people will believe the lie. There really is no counter speech to some falsehoods, some impersonations.

Dr. Wenstrup. And if we get a chance, I would love to hear back from you on the notion of extradition laws and other punitive measures.

Thank you. I yield back. My time is up.

RPTR PANGBURN

EDTR HOFSTAD

[10:02 a.m.]

The Chairman. Ms. Sewell?

Ms. Sewell. Thank you, Mr. Chairman.

So, Dr. Doermann, you didn't really answer my colleague's question about how far away are we from actually being able to detect defects. So I know that at DARPA you were working on that. Where are we either commercially or by government or researchers to, you know, technologically be able to detect deepfakes?

Mr. Doermann. So deepfakes is typically referred to as a, you know, particular technology that there is certain software out there for doing that. It is not a general concept.

So the deepfakes was actually -- the initial paper that was published that gave rise to this technology came after the start of the MediFor program. And we did adapt to start looking at those things.

There are point solutions out there today that deepfakes coming from these particular softwares can be detected. And, again --

Ms. Sewell. So do we have the technology to actually be able to digitally verify the videos or photographs, et cetera?

Mr. Doermann. The problem is doing it at scale. The problem is doing it at scale. If you give me a particular video, with high confidence, I can tell you whether this is a fake video. And I can also come back and say, okay, here are the videos or here are the images that went into it. Because typically it is --

Ms. Sewell. How long does that take? Is that a matter of an hour?

Thirty minutes?

Mr. Doermann. With the right hardware and things, you know, it can be done with a constant delay. So, yes, 15 minutes, 20 minutes.

Ms. Sewell. So, in advance of the 2020 elections, what can campaigns, political parties, candidates do to prepare for the possibility of deepfake content?

Mr. Watts?

Professor Citron?

Mr. Watts. So one thing, I think, even here on Capitol Hill and with political parties, is urge the social media industry to work together to create unified standards.

So part of the problem with all of these incidents is, if you are a manipulator, domestic or international, and you are making deepfakes, you are going to go whatever platform allows you to post anything from inauthentic accounts. So, if they can't share across accounts, it is like a cancer. They go to wherever the weak point is, and it spreads throughout the system, to the point where it really can't be policed, even if Facebook or Google or Twitter do a good job.

So I think one thing is really pressuring the social media industry to work together. And that goes for extremism, disinformation, political smear campaigns, all the things across. What is the standard for policing?

And then I think the other thing is having rapid responses to deal with this stuff. Any sort of lag -- as much as defense is not the best way, any sort of lag in terms of a response just allows that conspiracy to grow. So the quicker you get out on it.

Then mainstream media outlets can also work to help refute. Other politicians, other elected officials can help you do that refutation.

Ms. Sewell. Professor, what would you suggest that political parties and candidates do?

Ms. Citron. I think candidates should have clear policies about deepfakes and a commitment not to use them, not to spread them.

And then also to have relationships -- establish, early on, relationships with social media companies, so that when a candidate can say, you know what, I wasn't there, I wasn't doing or saying that at that particular time, to have immediate entree to folks at the content moderation, you know, whoever it is at Facebook, whoever it is at Twitter, Microsoft, whoever it is, that they have immediate sort of rapid response teams.

Ms. Sewell. How do we even begin to tackle this sort of liar dividend --

Ms. Citron. Right. Oh, I love that.

Ms. Sewell. -- in which --

Ms. Citron. You are using my phrase with Bobby Chesney.

Ms. Sewell. I know.

Ms. Citron. Thank you.

Ms. Sewell. -- in which politicians who may be recorded committing an illegal act can deny the truth by claiming that the recording is a deepfake? What do you suggest we do about that conundrum?

Ms. Citron. Oh, Congresswoman, I love this. Twice we have gotten some play for the liar's dividend, which Bobby Chesney and I conceived in our California Law Review piece.

And what most worries us is that, in an environment of pervasive deepfakes, where, you know, we have culturated people not to believe their eyes and ears, that the wrongdoer can seize on the fact that there isn't -- can take a genuine recording of mischief and say, "That is not me. That is a deepfake."

And so I think part of our -- so I have a twofold answer. Part of it is education. Part of our robust education that we have to have with the public is telling them about

the phenomenon of the liar's dividend, right?

So it is not that we shouldn't educate people. You know, so often the response to Bobby and I is, well, do we give up? Right? The liar's dividend -- do we stop educating? And our response is, absolutely not, that it must be a robust part of the learning curve, is to say, look, we know that wrongdoers are going to seize on the deepfake phenomenon to escape reality, and we can't let them do that either.

And so we have to somewhere get in the middle from completely believing everything our eyes and ears tell us to being skeptical without being nihilistic. Right? Because we do have a marketplace of ideas that is potentially functioning, but when what we are saying is not what we are saying -- we don't want to get into that space where we have a nonfunctioning marketplace of ideas.

Ms. Sewell. Thank you.

Thank you.

The Chairman. Mr. Stewart.

Mr. Stewart. Thank you, Chairman.

And to the witnesses, thank you for being here. It has been a helpful panel, although I have to say that I am a little bit concerned with some of your suggestions. I think, although in an ideal world they would be helpful, in the real world we live in I am afraid some of them are nearly impossible to implement. And some of them have some troubling aspects in themselves, in the sense that it is kind of like fact-checkers who aren't really fact-checkers; they insert their opinion. And this is just reality, a challenge we have before us.

Sitting on the Intel Committee, I am often asked, you know, in just conversations and casual discussions, you know, what do I think is the greatest threat facing the world. And a couple years ago, I answered that and I said, without thinking -- I nearly blurted it

out. I said, I think it is that no one knows what is real anymore.

And as I was driving home that evening, I started thinking on that, and I realized that is true. I think that is the greatest threat facing our Nation, where people just don't accept basic truths and basic falsehoods any longer, partly because of their own interests or partly because they just don't understand what is really true.

And it is not just deepfakes, by the way. RT television, for example, is extraordinarily good at propaganda that many people just think is perfectly legitimate and perfectly real. "Fake news," a term that we have all, unfortunately, become very familiar with.

The manipulation, as Mr. Himes has indicated and we can't discuss here, but manipulation of intelligence products is extraordinarily troubling to me.

And we live in a world where black is white and white is black. And I could show you evidence that white is black, and a lot of people would believe me that white is black.

And I just think, for us to lose that -- and, by the way, I think we can control governments. I think we can control, to a certain extent, legitimate businesses. But we can't control everyone. And this is going to be so pervasive and so available that virtually anyone could create this. And it is easy to control the U.S. Government and say, well, you can't create it, you can't use it for political manipulations or whatever it might be. But you can't control the other 6 billion people on the Earth.

And that is my concern, just the pure volume of it. It is like trying to monitor every bumblebee that is flying around America.

And the last thing, and then I will get to my questions, it goes both ways. And this is my concern as well. We could create the impression that a lie is real, but we could also say that something real is a lie, you know?

To use some of your examples, a politician caught in a bribe -- which, by the way,

politicians do much worse things than that. That is so 1970s. But let's go with that example. A politician is caught in a bribe, and it could be actually true, and he would then say, "No, no. It is just a deepfake. That is not real." And so you lose the credibility in both ways.

Which now brings me to my question. The first is -- and I will ask them both and get you to respond. With the potential for so much harm, should we have -- the algorithms that create deepfakes, should they be open-source?

And if the answer is, no, we have to do that right now, we can't wait for 2 or 3 years to do that because they will already be pervasive throughout the world, then the second question is -- and this is almost rhetorical, but I would love your answers or thoughts on this -- how do we prepare people to live in a world of deception? How do we prepare people to live in a world where they just generally may not know what is real or not?

Anyone who would jump on those two. Should the algorithms be open-source or should we control that?

Mr. Doermann. So, yeah, I will address the first one.

We made a conscious decision to make the MediFor program open. You will see, even a week and a half from now in Long Beach, at the Computer Vision and Pattern Recognition Conference, there will be a workshop there that is dealing with this.

Even though there is potential for our adversaries learning from these things, they are going to learn any way. We need to get this type of stuff out there. We need to get it into the hands of users. There are companies out there that are starting to take these types of things.

So I absolutely think that these types of things need to be open-source. There is nothing -- it is the same technology that is being used in terms of deep learning to create

this type of content.

Mr. Stewart. And I just want to say -- and you are saying that it should be open-source primarily because they will get access to it anyway. Is that the essence of your response?

Mr. Doermann. Well, and people need to be able to use it. The more we can use it to educate the community, educate people, give people the tools so that they can make the choices for themselves, that is what we are looking for.

Mr. Stewart. All right. I will accept that, although with some hesitation, but primarily the first part of the answer, and that is I think they would get it anyway.

What about suggestions on how we prepare people to live in a world that just is so steeped in deception? Have you -- I am sure you have thought through that. And we have 10 seconds to answer.

Ms. Citron. Oh, golly. Okay.

So, when Justice Oliver Wendell Holmes came up with the notion of the marketplace of ideas, he was a cynic. He wasn't suggesting that truth would always win out, and he worried about humanity.

But the broader endeavor is at the foundation of our democracy is that we can have a set of accepted truths so we can have real, meaningful policy conversations. We can't give up on the project.

Mr. Stewart. Well, I agree with you, and that is our hope. But, as I said earlier, that foundation of accepted truths is very shaky at this moment.

Thank you, Chairman.

The Chairman. Thank you.

Mr. Carson.

Mr. Carson. Thank you, Chairman Schiff.

In an era with prevailing distrust of journalists and the media, do deepfakes risk aggravating this kind of distrust?

Mr. Clark. So, prior to working in AI, I was a professional journalist for 7 or 8 years and finished up working at Bloomberg and Business Week, so I speak from some experience.

Yes, I think this is a very, very, very severe and potentially undercovered threat. Because when you write a story that people don't like, they try and attack you as the author of it, or they try and attack the integrity of the institution. And this makes it trivial to do that and to produce stuff that can convince people that you were not being factually accurate.

Ms. Citron. So, yes, not only do we see the journalists themselves, like Ms. Ayyub, being attacked, but I think what is so corrosive is the notion that the media is going to sit on real evidence for fear that it is a fake.

And we have certainly already seen, you know, sort of stings with media organizations. And now they have to be wary of stings with deepfakes that are really tough to debunk without some legwork, with journalistic effort.

And so the corrosive effort, what Bobby Chesney and I call trust decay, it affects not only politicians and our view of civic and political institutions but everything -- right? -- and, centrally so, journalism and the media.

Mr. Watts. If I could just add to that, over time, if an information consumer does not know what to believe, if they can't tell fact from fiction, then they will either believe everything or they will believe nothing at all. If they believe nothing at all, it leads to long-term apathy, and that is destructive for the United States.

I think you could look to Russia as an example of what has happened internally to the Russian people and how the Russian Government has used the firehose-of-falsehoods

approach, that if you can't believe anything, you just give up and surrender.

The consequences for democracy are political participation, long-term apathy, disappointment in officials that anything can be achieved, not wanting to show up and do things like register for the draft or show up as an all-volunteer force. I would tell you, that would be one I would look at over the next 10 to 15 years.

So I think that is the long-term corrosive effect. And if you look at Russia's long-term doctrine of subversion, that is what they are after. They just are much more patient than we are and willing to wait decades for it to come to fruition.

Mr. Carson. In addition to that, will technology solutions for authentication be available for even sufficient amounts for journalists or media organizations or fact-checkers to keep up with even validating a piece of media before reporting on it? Like Chairman Schiff, you know, crowdsurfing at South by Southwest or premiering his own Netflix special, how can you verify those things as journalists?

Mr. Doermann. I think it is important to have these tools out there for them. We are not at the case now -- as I said, we have point solutions. We don't have a general solution. We don't have a gatekeeper that can be automated completely.

This is a cat-and-mouse game. As things get better for our, you know, being able to deceive visually, they are going to get better and they are going to move on to covering up their trace evidence.

But I think the tools can be put in the hands, and they should be. We had situations where there are imbedded reporters, where somebody comes up to them with something on a cell phone and shows an atrocity. And, you know, they need those tools. They have to know whether to report on that or not.

So these manipulations, even before the automated piece of deepfakes, people were doing these types of things, and it is a major concern.

Mr. Carson. It could even evolve into some kind of new scam where you have someone with a piece of information selling it to TMZ or even a credible, so-called credible, media outlet, you know, scamming for $50,000, and the piece is, like, fake, you know?

Mr. Doermann. It is possible, sure.

Mr. Carson. Yeah, sure.

Mr. Watts. If I could add one dimension to this, though, is that how lucky we are to have a very engaged public in terms of actually rebutting things that are false and challenging them. You know, it is not just journalists that are doing it; it is also the public that will challenge back and forth.

One of the dangers that we don't think about, though, is in information environments where authoritarians control and eliminate all rebuttals. That can have a very significant backlash to us, which is why I would like to see widespread proliferation of authentication, not just for here in the United States but on the other side of the world where a regime controls all the information flow and can suppress reality.

Mr. Clark. It is worth stating, fact-checking is expensive and time-intensive, and the number of news organizations on the planet who are doing well in economic terms is sort of dwindling over time. And so I think that, if we were to go down this path, you need to find a way to fund that, because they, of their own volition, because of the economics, they are not going to naturally adopt this stuff, other than a few small, trusted institutions.

Mr. Carson. So it becomes incredibly difficult to remain a credible news source when you are having to pay to fact-check constantly.

Mr. Clark. Yes.

Mr. Carson. Okay. Thank you.

Chairman, I yield back.

The Chairman. Mr. Crawford.

Mr. Crawford. Thank you, Mr. Chairman.

Well, we have come a long way since Milli Vanilli, haven't we?

Just in the time that we have been here, I pulled up a video that was recently posted. Two British artists teamed with an Israeli company, if I get the name right here, Canny AI. I don't know if you are familiar with them. They created a video of Mark Zuckerberg saying, among other things, he could control the future. And they posted that on Facebook specifically to challenge Facebook. And then Zuckerberg has responded by saying he is not going to take that down.

I just wonder if you all could comment on that. What do you think this is about? And do you think it is a wise decision for Zuckerberg to not take it down, given what we have talked about?

And I will start with you, Professor Citron.

Ms. Citron. So I think that is a perfect example where, given the context, that that is satire and parody. That is really healthy for conversation.

And all of these questions are hard, right? Of course, our default presumption, as we approach speech online, is from a First Amendment perspective, which is we want to keep government out of calling balls and strikes when it comes to ideas in the marketplace. But private companies can make those kinds of choices, and they have an incredible amount of power. They also have free -- without any liability.

And I think they made the right choice to keep up the -- it was a conversation about, essentially, the cheapfake of Nancy Pelosi. It seemed to be a conversation about the choices that they made and what does that mean for a society. So it was incredibly productive, I think.

Mr. Clark. It seems correct in this instance. But all of these companies are kind of groping in the dark when it comes to what policies they need overall, because it is a really, really, really hard problem.

And so I think what would be helpful is to have a way for them to share policies across multiple companies and to sort of seek standardization. Because these judgment calls are very qualitative in nature, and they are going to become more numerous over time.

Mr. Watts. I would just add, though, that, while that is a comparison to what happened with, you know, the Congresswoman maybe being inebriated, a video -- they were trying to essentially duplicate that -- this does point out the idea of context, right? Part of that video, it spread for one purpose only, which was to challenge this rule so we would sort of discuss it in this forum, but no one really believes Mark Zuckerberg can control the future, because he surely wouldn't want to show up here to testify or anywhere else or be in the quagmire he is in. How do you know that?

I am trying to make a very serious point about context, which is, whenever virality spikes, that is where the assessment, I think, needs to come in terms of triage, and that assessment is when it goes into human curation. So that human curation -- okay, we see 4,000 shares in 10 minutes, now we see 16,000 shares over 15 minutes. That is when it should go. And then we look at labeling, we look at context, how do we inform the public so they make good decisions around it.

We had a parallel to this in the analog era. I, if I was a kid, would show up at a newsstand, and it would say, "Aliens Landed at Area 51." I would ask my mother, you know, friends, or family, where does this come from? And they would say, oh, that source is just putting out information for entertainment, that didn't really happen.

We need to help the consumer make a better decision around that. So I like it

that Facebook has been consistent in terms of their enforcement. And I am also not going to say that they should never change what those terms are. I think they are looking to here, to Capitol Hill, to figure out, what is it that we want to be policed? What does Europe want to be policed? I think they would like to hear from legislators about what falls inside those parameters.

The one thing that I do really like that they are doing is inauthentic account creation and inauthentic content generation. They are enforcing that, and they have increased it. And I think that is really, really good in terms of how they have scaled that up. It is not perfect, but it is better.

Mr. Crawford. Let me ask you this. Is there a particular company or is there a particular region or particular nation that is especially adept at this technology, that is developing it at a quicker rate or whatever?

Mr. Clark. It is distributed along the lines you would expect of prominent research centers in America and China and Europe. So it is distributed. Wherever you have good AI technologists, you have the capability to create this stuff, which makes it very challenging.

Mr. Crawford. At some point this will be available off the shelf, though, right? Folks at home will be able to access, it as all technology --

Mr. Clark. It already is.

Mr. Doermann. Absolutely. That is one of the big differences. You used to have to go out and buy Photoshop or, you know, have some of these desktop editors. Now, you know, a high school student with a good computer -- and if they are a gamer, they already have a good GPU card -- can download this, can download data, and train this type of thing overnight with software that is open and freely available.

So it is not something that you have to be an AI expert to run. A novice can run

these types of things.

Mr. Crawford. Thank you.

I yield back.

The Chairman. Mr. Quigley.

Mr. Quigley. Thank you, Mr. Chairman.

Thank you for your participation.

Following up on those points, the themes here: getting easier to do, the quality is getting better, getting harder to detect. The examples we talk about as victims -- democracy, elected officials, corporations, this horrible attack on a journalist. But what about a small business with limited resources? What about individuals who are victims of, as the example you gave, Professor, revenge porn, for example?

And, Doctor, you talked about the scale and widespread authentication. What capabilities might exist as we go forward either on social media platforms, law enforcement, or for individuals themselves to deal with this detection issue?

Mr. Doermann. Well, you know, I envision some time where, you know, there is a button on every social media piece or every time you get even a text message with a video attached to it that you can hit, it goes off, it gathers information, not necessarily totally automated -- if it has been vetted by one of many other organizations, if you can identify where it came from -- so that the individual can make those decisions.

The problem is that a lot of these types of technologies exist in the labs, in research, in different organizations; they are not shared, and they are not implemented at scale.

So if I want to go out and test a picture -- there was a very interesting picture before a tornado up in Maryland a couple of weeks ago. It looked surreal, and I immediately thought, oh, that must have been somewhere else, you know, somewhere in

the Midwest, you know, years ago. So I did a search. There is a reverse image search that you can do. And after, you know, doing some research, I found that it indeed was real and it was practically in my backyard.

But not everybody has those types of capabilities. Not everybody thinks to do that type of thing. I know that I have relatives that, you know, just do this and they see something and they want to share it. And so I think the education piece and getting these tools at scale is what we need to work towards.

Ms. Citron. But the key is, even with detection, for the everyday person who has a deepfake sex video in the Google search of their name prominently featured, and a platform refuses to take it down, it is their CV, meaning it is part of what everyone sees about them. And so it is incredibly destructive.

And the same is probably true for the small business; can't afford reputation.com. If there is a deepfake that really casts asunder under their business model, they may not be able to have it removed even though it is false and it is an impersonation and even if it is defamation. We know that the law moves really slowly if they brought a defamation suit, assuming they could find who the creator is.

So we are going to have this -- we are in this liminal period where -- and it may last years -- where individuals will suffer. And it is incredibly hard to talk to victims, because there is so little that I can force anyone to do. And we are going to see a lot of suffering.

Mr. Quigley. And the issues that we just talked about, are you trying to tackle those with your model laws that you were talking about?

Ms. Citron. Yes. So I am the vice president of the Cyber Civil Rights Initiative, and we have been working with lawmakers around the country, both at the State level and the Federal level, both in terms of nonconsensual pornography and now to think about how we might really carefully and narrowly craft a law that would ban deepfakes or

manufactured videos that are essentially impersonations that amount to criminal defamation.

So I think we have work ahead of us at CCRI and laws around the country. It can be tackled, but it is going to have a really modest impact, because the law moves slowly.

Mr. Quigley. And when you are doing this, are you talking to the local and State law enforcement agencies?

Ms. Citron. Yes. So, in my work on cyberstalking, I wrote a book called "Hate Crimes in Cyberspace," which was about the phenomenon of cyberstalking and how hard it is to teach local law enforcement both about the technology and the laws themselves. You know, they are great at street crimes, but when you talk to them about online crimes, even though there are offline components, they say, I don't really know where to begin, I don't know how to get a warrant for an online service provider to get an IP address to go to the ISP.

So we do have some education. I know Congresswoman Clark has called for funding, some training of local law enforcement on the question of cyberstalking, both as a technical matter and then as to law. And I would love to see that not only with regard to cyberstalking and threats but more broadly.

Mr. Quigley. Thank you all.

The Chairman. Mr. Hurd.

Mr. Hurd. Thank you, Chairman.

I am going to try to do something that is probably impossible in the next 5 minutes and touch on and get you all's perspective on four areas: the ability to detect -- Mr. Doermann, you touched on authentication as a strategy for this. How do we handle -- how do we develop a strategy, in a narrow national security sense, to counter disinformation, and who should be doing that. And then, broadly, education.

And my first question is probably to you, Mr. Clark and Dr. Doermann. Can you talk to us about the ability to detect and the forensics? Like, is there ability to do a pixel-by-pixel analysis? A bit-by-bit analysis? Are there others areas of basic research that we should be focusing on in order for us to help with an ability to detect?

Mr. Doermann. Well, the approach that is being taken in the community is one of a comprehensive view. So, yes, there are pixel types of applications, not necessarily pixel by pixel, but the metadata that you get on an image, you know what compression algorithms that were there. You know that there are residual information left if you take an image, you modify it, and you recompress it.

So, at the digital level, that is where a majority of the work is being done. And that is the --

Mr. Hurd. And how easy is that now, and who should potentially be doing that?

Mr. Doermann. Well, the government is putting a lot of money into this piece. As I said, there are a lot more manipulators than there are detectors. So I would hope that behind closed doors the social media sites and, you know, the YouTubes of the world are looking into this type of application, but I am not sure.

Mr. Hurd. And is the ability to understand the various metadata or even getting to a point where we can do pixel-by-pixel, you know, exploration en masse, is that going to help us to a point where we can do real authentication? So, anytime you put a video up or a picture up, there is a green checkmark, you know, is that --

Mr. Doermann. Personally, I don't like to use the word "authentication" because, as we know, absolutely everything that goes up online is modified in some way, whether it is cropped or, you know, there is a color histogram distribution --

Mr. Hurd. What word would you use?

Mr. Doermann. -- adjustment. Well, we like to use that things have been

modified.

But it is a scale. So if there is the modification of intent, if you put a flower in a picture next to someone, that has a very different effect than if you replace somebody's face in a picture.

And so this discussion, this attribution piece and the actual report that says this is exactly what was done was a big part of the MediFor program as well.

Mr. Hurd. Uh-huh. So the closest you are going to get is to say, all of these things happened to this image, and, therefore, the user would have to be the one to make the decision on whether this is credible or not.

Mr. Doermann. Yes. And even in an automated way, if you are taking an image and you are the FBI and you are going to court, even if you did change one pixel, you lose the credibility.

Mr. Hurd. Gotcha.

Mr. Doermann. But if you are FBI and you are doing an investigation and you have a very compressed, grainy surveillance video, it still might give you information and you believe it.

Mr. Hurd. Ms. Citron and then maybe Mr. Watts, you know, disinformation is a subsection of covert action. Covert action and counter-covert action is the responsibility of the Central Intelligence Agency. Yet the Central Intelligence Agency, because of the National Security Act of 1947, can't do covert action in the United States of America. Very hard to do it in English.

How should we be looking at a government strategy to deal with disinformation, especially in the context of national security?

Or somebody else more appropriate to start with that.

Mr. Watts. I think it is two parts.

I would encourage the social media industry and the platforms to focus on methods: Who is doing deepfakes, digital forgeries? Who is doing computational propaganda? Can we have a listing of those? They are not always nefarious. But then we know who the people are that are building the equipment. This is essentially the weapons that are being used.

And I would encourage the government, then, to focus on actors. So this is in the case of -- the CIA oversees DHS in terms of protecting the homeland. Who are -- you know, State Department, which used to have the U.S. Information Agency, would be out there outing and overtly going after those actors that are doing the manipulation. I feel like we are still, after several years now, really slow to do this. And they are the only ones that can figure it out.

When I have worked with social media teams and we spot actors that we believe are doing things, we sometimes have to wait years for the government to go, yes, here is the Mueller report and it labels the Internet Research Agency. But that had already been out in the news.

So the more rapidly the government can do that, the more the public can help, the more the social media companies know what to take down -- because that attribution really only comes down to the U.S. Government. They are the only ones with the tools, really, that can do that.

Mr. Hurd. Good copy. Thank you.

Chairman, I yield back.

The Chairman. Mr. Heck.

Mr. Heck. Thank you, Mr. Chairman.

First of all, Professor Citron, I want to make sure that I understood correctly. If something like happened to that reporter in India had happened in America, did I

understand correctly that that would not constitute a crime per se?

Ms. Citron. It might be understood as cyberstalking, which is a crime under Federal and most States' laws. The problem is it was sort of like death by a thousand cuts. And to constitute cyberstalking, you need a course of conduct, a persistent, repetition by the same person. So it is --

Mr. Heck. So if it were the first time --

Ms. Citron. And what happens is, it is like a cyber mob coming together. So one person puts up the photo, a screenshot; another person puts up the home address; yet another person puts up "I am available," just all it says is --

Mr. Heck. Understood.

Ms. Citron. -- "I am available" with a screenshot --

Mr. Heck. So the person who originated it, under current law, would likely not be subject to criminal prosecution.

Ms. Citron. Right.

Mr. Heck. Did I also understand you to say that, even if it were, it would have modest impact?

Ms. Citron. What I said was, if we had criminal laws that combated the sort of deepfake phenomenon really tailored to falsehoods/impersonations that create cognizable harm, I think law is really important. It is not that it -- it is modest in the overall impact, because we need a partnership with technologists, with --

Mr. Heck. Okay. I am sorry. I want to move on. But I also cannot help but have this terrible flash of Dante's Inferno: "Abandon hope, all ye who enter here."

Whose job should it be to label? That wasn't clear. I kind of thought it might be the media platform companies.

Ms. Citron. I think it would be the creator -- that we could, much as we do in the

campaign finance space, where we say there are certain disclosure rules, that we say if it is a political ad, you have to own it. It --

Mr. Heck. So, if it is a foreign originator, how is it that we have any jurisdictional reach?

Ms. Citron. We don't.

Mr. Heck. I mean --

Ms. Citron. Right.

Mr. Heck. -- there are no boundaries, right? And so, as a matter of practical fact, even if it is created in America, transmitted to a foreign person, and then retransmitted, we have no means of enforcement.

Ms. Citron. Right.

Mr. Heck. So labeling, in and of itself --

Ms. Citron. And that might be --

Mr. Heck. See remarks above about Dante's Inferno.

Ms. Citron. But, look, we have social media platforms. If they had some responsibility, they may -- and if, indeed -- and I am pretty skeptical about whether we are going to get there in the near future about the technology of detection. But assuming that is possible, then a reasonable practice could be disclosure, saying: This is a fake. Do with it what you will.

Mr. Heck. So we actually have, as it were, a comparable truth verification mechanism currently, Snopes. And yet a member of my family, who shall go unnamed, immediate family, once posted how outrageous it was and how the Constitution ought to be amended because Members of Congress can retire after one term, immediately collect full pension benefits -- every Member up here has heard this -- have healthcare free for life, and their children go to college for free.

Not one word, not one letter of that assertion is true, which could have easily been verified if they had gone to Snopes. They didn't. And even if they did, in a political context, the truth is, the person who is perpetuating that may have a political agenda such that they also, in a parallel fashion, engage in ad hominem, as it were, attacks against the reliability of Snopes.

So see remarks above. Abandon hope, all ye who enter here.

I don't have much time left, but I am really interested in Mr. Himes' getting at the issue of political speech and the First Amendment. You mentioned that we are protected against being impersonated, but it is not clear to me how we square case law, which has created a very high --

Ms. Citron. Yeah.

Mr. Heck. -- barrier --

Ms. Citron. And it is incredibly important to recognize that everything you have just described is totally protected speech. The United States has made clear in, a case called the United States v. Alvarez, a plurality and concurrences of the Court said that, look, we protect falsehoods, that we are going to ensure it enjoys First Amendment protection, because, as Justice Kennedy explained, it reawakens the conscience. It has us engage in counterspeech and sort of recommit to citizenship.

But there are times, as the Court made clear, that when falsehoods create certain kinds of cognizable harm, that we can and should regulate. And that includes defamation, even of public officials if said with actual malice, that you know or you are reckless as to the truth of the matter asserted.

So there are true threats, incitement -- there are 21 crimes made up of speech. We can regulate certain words and images if it falls in one of those categories or if we can meet strict scrutiny.

So, yes, the presumption is that it is protected speech if it is a falsehood. But falsehood that causes cognizable harm, the Court has explicitly said -- actually, it is the entire Court -- has said that that is a space that we allow regulation.

Mr. Heck. Thank you.

The Chairman. Mr. Welch.

Mr. Welch. Thank you very much.

You know, this is very helpful. There are different categories, and we are all trying to get our arms around them. There is the question of the First Amendment, which Mr. Heck and Mr. Himes were talking about. There is the question of foreign interference. And there is the question of economic harm, reputational harm. And we are all learning as we go on this.

But what I have heard you be describing is, essentially the whole world of publishing is upside-down. It doesn't exist like it did prior to the internet. So the question is whether we want to get back to some of the principles that applied pre-social-media. It is not like those principles necessarily have to be abandoned. They have to be applied. And they would apply in different ways for each of those different categories.

So I just want to ask each of you whether we should get back to the requirement of an editorial function and a reasonable standard of care by treating these platforms as publishers.

And I know, Ms. Citron, you said yes. I would be interested in what the others say.

Just "yes" or "no" on that.

Mr. Doermann. So, working with a number of people in this area, I think, you know, the horse has sort of left the gate on this. I don't think we are going to be able to

get back to that type of editorial --

Mr. Welch. Well, what about with that statutory change that Ms. Citron was proposing?

Let me just go on a little bit, because who has the duty --

Ms. Citron. May I be clear for one second?

Mr. Welch. Yeah.

Ms. Citron. It wasn't that I was suggesting that social media platforms be understood as publishers, strictly liable, but, rather, that we condition their immunity on reasonable practices.

Mr. Welch. Right.

Ms. Citron. Those reasonable practices may be the content moderation practices they use right now.

So I am going to disagree about calling them publishers who would be strictly liable for defamation. That is not what I am suggesting at all.

Mr. Welch. I see. Okay. Thank you for that clarification. But that seems to be one fundamental question that we would have to ask, because that would be a legislative action.

Mr. Clark?

Mr. Clark. I think you have a whack-a-mole issue here where people, sort of, online go and talk, and they compose their own platforms very, very quickly, and they compose platforms to evade rules that we put against platforms doing certain types of things.

So I do agree with Dr. David Doermann here that it is very difficult to contemplate controlling speech in this way, because I think the habit of the entire culture of people has changed. So people --

Mr. Welch. What about on this question of somebody going online and putting up a fake video that destroys an IPO? I mean, who has the duty of care with respect to allowing that to be stated on their platform? Nobody has it?

Mr. Clark. I think we can authenticate content and users, and I think that you can make users culpable for certain types of content that they post. And --

Mr. Welch. So who would be liable in the case of that false statement about an IPO that destroyed its value?

Mr. Clark. I will defer to the lawyer.

Ms. Citron. The speaker. The speaker. The creator of the deepfake. And so long as, as Mr. Clark suggested, the platform had reasonable practices of authentication and ex-post moderation practices --

Mr. Welch. But does the platform, under current law, have any duty?

Ms. Citron. They have no liability under section 230 of the Decency Act.

Mr. Welch. Right. So that seems like a very direct question.

Ms. Citron. Right.

Mr. Welch. One of the other issues that is debated -- there is a different point of view, often, between Republicans and Democrats about bias and what goes on the platforms. So there would have to be some standard that wasn't seen as tilting the playing field for Republicans or Democrats. Is that possible to do? And was that something that was true pre-social-media, in the days of --

Ms. Citron. You are the journalist. What do you think, Jack?

Mr. Clark. Well, I was going to say that, for standards, we can actually use technology a bit here to create technological standards for making a judgment call as to whether something is or is not faked or synthetic. And that might take the political aspects out of this, if you have open standards developed primarily by academia, the

companies chime in on, and it is auditable by scientists who kick the tires and provide assurance. That seems reasonable.

Mr. Welch. Okay.

Thank you all very much.

My time is up. I yield back.

The Chairman. Ms. Demings.

Mrs. Demings. Thank you so much, Mr. Chairman.

And thank you to all of you for being here.

You know, this conversation this morning has been pretty disturbing and actually quite scary. You know, when I think about it, the internet is the new weapon of choice.

And as I listen to the testimony and the questions here, you know, as we think about an individual who goes out and violates laws or creates harm, they would be held accountable. I believe that any individual or entity that bullies or stalks or creates harm or a public safety becomes a public safety risk -- any entity that creates an environment for those things to happen should be held accountable as well.

And, you know, when I think about those around the world who are not our allies, they want to create chaos in this country. And what a wonderful way, an easy way to be able to do that.

The problem that, you know -- of course, the fake information is a problem. But the other problem is, it creates an environment where good people no longer believe the good guys. And, boy, are we seeing that in our country right now. That is a major problem. Our institutions that we have grown to depend on and believe in are no longer being believed. And that can create total chaos.

Back to Mr. Heck's statement about, say, for example, a fake video is created in America but then transmitted to another country. Could not the act of transmitting, the

simple act of transmitting that video be the violation?

Because I know there has been a lot of discussion about there are no boundaries, how do you hold someone in a foreign land accountable or -- but I would love to hear your thoughts on that.

Ms. Citron. So I think there are two pieces to that. There is the sort of procedural, jurisdictional question of whether it is constitutional to haul them into your court, personal jurisdiction. And then there is the extradition question, which I am going to rely on, I think, Mr. Watts for that. But --

Mrs. Demings. If you are in America and you transmit a video that creates a public safety concern or a national security risk, could not the very act of transmitting it from America be the violation, as we talk about policy?

Ms. Citron. So you are asking -- so it is directed outside the United States?

Mrs. Demings. Uh-huh.

Ms. Citron. So you are in the United States and you direct --

Mrs. Demings. Directed outside. Transmitted from Florida.

Ms. Citron. Okay. So that is a different question than I thought you were asking, because when we -- under the 14th Amendment, how we think about personal jurisdiction is that we say, if you are aiming your activity and you are aiming a tort to another State and you are doing it purposefully, you are availing yourself of that State, we can haul you into court so long as there is a long-arm statute.

But now -- you can tell I teach civil procedure. But now you have confused me a little, because then the question is -- when it is an American directing harmful activity abroad, I would imagine that, then, is contingent on that country's jurisdictional rules and our extradition arrangements and treaties with them.

So, Mr. Watts, do you want to take it from there?

Mr. Watts. Yeah. I mean, I am not a lawyer, and I try to avoid them, but I would generally say that there is no specific provision around transmitting that abroad. I think it comes down to whatever country it is that is affected by it, if it breaks their laws, and then if they have an extradition relationship with the United States.

That is probably not worked out. I am not sure if it has ever been executed. It could have and I am not aware of it. But it is something that needs to be addressed.

Because what has been very clear over the last 4 years is there is no physical boundary in these communities, in these disinformation networks online. And, oftentimes, the smartest manipulators out there -- Russia, China, Iran -- they actually look to enlist people in foreign lands to make the content look more authentic --

Mrs. Demings. Exactly.

Mr. Watts. -- and they are setting people up. Sometimes they are aware of it, and sometimes they are not aware of it. Those that are aware of it are doing it willingly.

And so, if you look at the Macron leaks, for example, which was another hacking attempt trying to drive an election, it was actually someone in North America that alerted the world to it and pointed the direction to it.

So I do think we need to figure out what those relationships are and how we would handle it, in terms of our own law enforcement environment. Because we are now going out to other countries and asking them to do that for us.

Mrs. Demings. Thank you.

And back to -- I know we have talked quite a bit, too, about the Intelligence Community and our national security entities, but could you talk just a little bit about, how should we task the Intelligence Community and our national security entities with assessing and forecasting future impacts of deep-state technology?

Mr. Watts. I think there are two parts. One, who are the purveyors and actors

that are going to use that? That is pretty straightforward. I mean, from the outside, even where I work, I can see a lot of that.

I think the part that might be missing from the government's perspective is where that technology is being developed. The number-one place I would have someone as a liaison in the U.S. Government right now is Tel Aviv. I mean, this has been a central hub of everything from cyber tools to influence tools, influence operations, both good and bad, you know, depending on what your perspective is, but that is a tech hub.

I feel like oftentimes when I talk to the government about that they are really well-informed about what deep nation-state actors are doing but oftentimes missing what the private sector has openly available in terms of AI and other tools that are out there.

Mrs. Demings. Thank you.

Yes, Mr. Clark?

Mr. Clark. Just quickly --

Mrs. Demings. Chairman, can he respond quickly?

The Chairman. Yes.

Mrs. Demings. Thank you.

Mr. Clark. Okay. Just quickly, to this point, it is worth repeating that the fundamental techniques and tools for this are published openly online, and we can easily compile quantitative metrics of a rate of improvement so we can do that forecasting.

So I agree with what Mr. Watts said, but it is easy to go and discover this information for ourselves.

Mrs. Demings. Thank you.

Mr. Chairman, I yield back.

The Chairman. Dr. Wenstrup.

Dr. Wenstrup. Thank you, Mr. Chairman.

And thank you, Mrs. Demings, because you addressed the question that we ran out of time on, on the extradition laws. I appreciate having the opportunity to hear from you on that and, you know, just getting to other punitive measures that we may be able to start talking about and thinking about.

You know, with the extradition laws, we might end up with a lot of people hanging out in other people's embassies for many, many years rather than being extradited. But, at the same time, while as a doctor I don't often find myself eager to engage with trial lawyers, that is probably where we need to head with this as people are harmed through all this.

So, you know, I am going through my mind, what kind of punitive measures? Certainly monetary would be included, because people end up, as we have pointed out, with huge monetary losses because of these fake stories. And what about prison time? I mean, I think we really need to consider being pretty tough on this if it is to be effective.

Mr. Watts. One thing I would add that Chairman Schiff brought up -- and I kind of ran out of time in the opening questions -- was about sanctions.

What we did see, if you look at the GRU indictment in particular, which I think is July of 2018, they are essentially being sanctioned or outed. And so are those companies in the February 2018 troll farm indictment.

That is very effective, but you could move down the chain of command such that hackers and influencers, propagandists don't want to be hired at those firms because they know the risk that they could be individually sanctioned. I think that could be an effective technique.

It seems like it would be hard to execute, but once we got good at it, I think it would be a great facet, which is, if you can turn down the employment to where the best

hackers and the best propagandists don't want to work with those authoritarian regimes, it could change the nature of things.

I think we could also look at those that are pushing out tools, both in terms of cyber and hacking tools that are being used for very malicious purposes and for influence techniques. You could actually go after those companies, which are oftentimes international. They are not necessarily tied to a nation-state.

And that would also send a downward pressure, you know, across the disinformation space. It would also send it more undercover and places like the dark web. But that is okay, because that plays to our strengths, which is, we have great intelligence collection capabilities at that end and we have good, sophisticated intelligence agencies.

Dr. Wenstrup. And now we would know where it is --

Mr. Watts. Right.

Dr. Wenstrup. -- more likely, but it would be a black market.

Mr. Watts. It changes the problem but, I think, to our advantage. We are moving in the right direction.

Dr. Wenstrup. Well, the other thing, too, is, you know, you mentioned sanctions, and that does make a lot of sense, especially if it is a country that there is no way you are going to get some type of extradition agreement in place, right?

Mr. Watts. Right. And I think that is the case with most of these locations, whether it is China, Iran, or Russia. Those are the three big ones. But it also would send a message out across the world that if you are pushing on us there are options that we have.

I do think that the time for offensive cyber is at hand. And General Nakasone, I think, has done some very good briefings recently about that, talking about the measures

they are taking.

If these foreign manipulators, makers of deepfakes that are working at troll farms, cyberhackers, knew that we were actually going to respond in a very aggressive way, they would move away, whether it is arrest and extradition, if it is sanctions individually, or even in terms of cyber response.

Dr. Wenstrup. Right now, there is not a whole lot of deterrence.

Mr. Watts. No. It has proliferated because we have not responded.

Dr. Wenstrup. Yeah. Thank you.

I yield back, unless someone else wants to comment, but I appreciate it. Thank you for the time.

The Chairman. Thank you.

I just had a few followup questions.

Can you talk a little bit about the -- oh, I am sorry. Mr. Castro.

Mr. Castro. Thank you, Chairman.

Professor Citron, first, I enjoyed your article with Bobby Chesney out of UT-Austin and had a chance to visit with him on some of these issues a few months back. And you mentioned the case about falsehoods. And I think this will be a monumental task for the legislative branch and then the judicial branch to grapple with how we treat deepfakes.

There is some speech, like hate speech and fighting words, that are not as protected, obviously, as political speech. And in making that determination, we have to figure out what the value is of the type of speech or expression.

So let me ask you, what is the value of a fake?

Ms. Citron. And just to add to that -- and thank you so much for reading our piece -- is the value to the listeners, right? So when we think about free speech theory,

it is the value to the autonomy rights of the speaker and for self-governance.

Mr. Castro. The creator in this case.

Ms. Citron. The creator but also, of course, the listeners.

Mr. Castro. Sure.

Ms. Citron. And so the value of the fake could be profound. It could be that the deepfake contributes to art. You know, "Star Wars," we had Carrie Fisher coming back. There is a lot of value in deepfakes. So I recognize what my co-panelists are suggesting.

But we do have guides in the law about falsehoods, impersonations that cause harm, whether it is defamation law or it is, you know, another kind of speech where we say fraud --

Mr. Castro. Right. So you think we may go down the road or the Court eventually may go down the road where certain speech, like hate speech, is not protected, obviously, the same way as political speech or even ordinary speech.

Ms. Citron. I think we are going to stay firm on hate speech. I have a feeling. And I think we --

Mr. Castro. No, but I mean for fakes, that there will be certain fakes that are treated differently than other fakes.

Ms. Citron. Right, depending on the context. You know, all of this is so contextual, so I don't think we can have a one-size-fits-all rule for deep synthetic video, even as to impersonations, because you could have satire and parody, which is really valuable and important.

At the same time, we have to bring content to the fore and say there are times when these falsehoods, the deepfake causes real harm -- real harm that is cognizable and real harm that either doesn't enjoy First Amendment protection or enjoys less rigorous

protection, and we can regulate it.

Mr. Castro. Let me follow up. I wanted to ask you all, you know, one of the big challenges that we had with the Russian interference, particularly what they put on Facebook and social media, is that it seemed as though the social media companies were unprepared for that, and there was no infrastructure for vetting or moderating those things.

So, you know, I mean, just my rough sketch -- obviously, you all have thought about this a lot longer, but I see that there is a creator who uses software, who then posts on social media, and then the traditional media picks it up and then further proliferates it into the bloodstream of American society.

So where in there do we construct that infrastructure for vetting and for moderating these deepfakes? Who is responsible at each of those levels?

Mr. Doermann. Well, you know, again, I am not the lawyer or the policymaker, but I think there is another piece to that puzzle. Somebody puts something up that is innocent, and it gets used by someone else for a different message. So, you know, this is almost not even the deepfakes problem but something that gets put out there and then gets twisted in a certain way somewhere down the line, way after --

Mr. Castro. I mean, there are a lot of people that don't realize, sometimes, that the Onion articles are actually satire.

Mr. Doermann. Exactly. That is a good example of that.

So I think we need these types of things at every level, that we need to be able to show the attribution of this information, how it progressed, and be able to make those decisions at every level.

Mr. Watts. I would add, I think that scenario is exactly what will happen going into future elections by our foreign adversaries, which will be to use as much organic

American content that suits their narrative and to amplify that and inject it back. That is a pretty standard disinformation approach.

And especially as false content is proliferating -- you know, more people are able to make it each year. They can make fake content. That means that it is more available for our adversaries to repurpose and reuse, which is the scenario that David just talked about.

So I think the social media companies need to work in terms of virality, what are their thresholds for doing assessments, how they do their content labeling, and then even a triage within that in terms of severity of impact. We know what some of those are, like mobilization to violence, talking to violence, but also in terms of effect to democracies and political institutions, things related to elections.

Right now, I would be very worried about someone making a fake video about electoral systems being out or broken down on election day 2020. We should already be building a battle drill, you know, a response plan about how we would handle that in the government, in the State governments and the DHS, as well as with the social media companies.

Mr. Castro. Thank you, Chairman.

I yield back.

The Chairman. Thank you.

I just wanted to ask a few followup questions.

I don't know if any of you know to date how many millions of views the doctored video of Speaker Pelosi has received, but I wonder if you have a sense of, if there are X million views of that video, how many of those millions will ultimately learn that that video was a fake and how many will be permanently misled?

And then, what is more, if you could comment on the phenomenon that, even if

you are later persuaded that what you have seen of the person is not true, I understand that psychologists will tell you that you never completely lose the lingering negative impression you have of the person.

So I wonder if you can comment on those two issues.

Mr. Clark. Fact-checks and clarifications tend not to travel nearly as far as the initial news. So we would expect the same to hold here, where for people who have seen the doctored video, a tiny minority will be aware that it was doctored, would be the assumption.

The Chairman. So the assumption will be, if you put this out, that a very small minority will actually learn that it is a fake, no matter how good you or the press do of putting that out there? Because the truth, in this case, that what you have seen is false, is not going to be as visually impacting, may not be visual at all, as seeing the video?

Mr. Clark. The way I would quickly put it is, if you care, you care about clarifications and fact-checks. But if you are just passively enjoying media, you enjoy media. And so the experience of the Speaker fake is, you enjoy or experience that media, and an absolute minority care about whether that is true beyond the entertainment value you extracted from it. Just as a general thing.

The Chairman. And, you know, what should -- you know, I don't know whether it is journalism or what now. What should teachers in schools be educating young people about these days about whether you can believe what you see?

You know, this gets to the liar's dividend. And, by the way, you know, in politics, there is a saying that the first time you hear an expression or an anecdote or a story, you make personal attribution. The second time, you say, "Somebody once said." And the third time, it is pretty much yours. So liar's dividend is now out there.

But, you know, how do we educate young people or not-so-young people about

how to perceive media now without encouraging them to distrust everything, in which case there is this liar's dividend?

RPTR FORADORI

EDTR CRYSTAL

[11:05 a.m.]

Ms. Citron. It is true that the more that what we are seeing, even if it is totally false, confirms our world view, social psychology studies show that we are just going to ignore it. So that we will believe the lie if it confirmed -- it is confirmation bias theory. So you are right. But it becomes incredibly hard for the fakery to be debunked because it is so visceral video, and because if it confirms your world view it is going to be -- it is really tough.

I guess that is the task of, as parents, as educators, as teachers, as we talk to our students, the critical -- 10 years ago I remember the critical thinking was about, how do we teach students how to do a Google search? And can they believe everything that is in the prominent search in whatever they are doing?

And we saw that, you know, you did a search for the term "Jew," what would come up first was a virulently anti-Semitic site called Jew Watch. And teachers struggled to explain to students that just because it is prominent doesn't mean it is real, doesn't mean that it is the authority.

And I think we are going to have the same struggle today, that, yes, we are going to teach them about deepfakes, and I think we have also got to teach them about the misuse of the phenomenon to avoid and escape responsibility.

The Chairman. I mean, the other challenge, too, is we have a White House that has popularized the term "fake" to describe lots of things that are real, in fact some of the most trusted news sources in the country.

So there is already an environment in which there is license to call things fake that

are true, but are critical. And it seems that that is a pretty fertile ground for the proliferation of information that is truly fake.

And we find ourselves, frankly, trying to find other words for it -- false, fraudulent -- because fake now has been so debased as a term people don't know really what you mean by it.

Ms. Citron. I think it is worth noting, too, that when President Trump referred to the "Access Hollywood" tape, he said: Well, that never really happened. The Holt interview: Oh, that wasn't right. We have already seen the liar's dividend happen in practice from the highest of the bully pulpits. So I think we have got a real problem on our hands.

Mr. Clark. I do think there is some optimism for tools. I have been involved in numerous arguments with friends where we have gone and checked where it is imperfect, something like Wikipedia. You end up using the information sources around you, and you can train people that there are certain sources you can go to settle an argument, as it were. And I think that we can develop such tools for some of this technology.

Mr. Doermann. I think that that is a great motivation for having this information up front. When Mr. Heck was saying that he had a family member that didn't know about going to Snopes, if that information was an attached to the video or the email or whatever ahead of time, they would have had access to it and they wouldn't have had to go search for it.

The Chairman. Well, I am just thinking of the applying in 2020 what we saw in 2016. And in 2016, among other things, the Russians mimicked Black Lives Matter to push out content to racially divide people. And you can imagine foreign bad actors, particularly Russia, dividing us by pushing out fake videos of police violence on people of

color. We have plenty of really authentic ones, but you could certainly push out videos that are enormously jarring and disruptive, and there even more so than seeing a false video of someone and still having that negative impression.

You can't unwind the consequence of what happens in the community. It is hard to imagine that not happening, because it is such low barriers to entry, and there will be such easy deniability.

Mr. Watts. If I could add. There is some good news in that social media -- if you watch Facebook's Newsroom, they are doing take-downs nearly every week now. So they have sped that up precipitously.

But we actually have the curriculum for evaluating information sources in the U.S. Government. I was trained on it at the FBI Academy, they have it at the Defense Intelligence Agency, Central Intelligence Agency, which is, how to evaluate information outlets, how do you evaluate expertise. They teach this, it is unclassified, there is no secret sort of course. But it is how you adapt that into the online space.

The audience I am most worried about is actually not young people in social media, it is the older generation who has come to this technology late, that doesn't really understand -- they understand the way newspapers are produced, where the outlet is coming from, who the authors are.

So I was with a group at New York City Media Lab and they actually had a group of student, it was, how do we help older generations new to social media or has less experience evaluating these sources? You can send them tips and cues. Do you know where this outlet is physically located at? That is one. Do you know who the author is, who the actual content provider is, or can the social media company tell you?

I think there are simple tools like that that we could develop or the social media companies could develop for all audiences, because it is not just for the young people.

Young people actually have more iterations oftentimes in information evaluation digitally than their parents do. They have actually done this at times more.

So I think in terms of thinking about approaches, it is about what is the generation, what are the platforms they are on? Do they understand that places like 8chan, which are known for extremism, is based in the Philippines, and that is not really in the United States in the sense of our ability to administer these things.

There are some simple tools I think we could do that are nothing more than widgets, public awareness campaigns, things that we can take from the government that we have already developed and really repackaged for different audiences in the United States.

The Chairman. Dr. Doermann, if I could, is the technology already at the stage where good AI can produce a video that is indistinguishable from real to people with the naked eye? In other words, could AI right now fool you if you don't have access to computer analysis of whether the video is authentic?

Mr. Doermann. Yes. I think there are examples out there that, taken out of context, that if they are sent out there and there is a story or a message with it, that people will believe it. And it is not just people that have that agenda. It was a video that was out there that showed a plane flying upside down, very realistic looking. And I think what people will need to do is get confirmation from other sources that something really happened.

So a video in isolation, if that is what you are talking about, a video in isolation, you are given this, asking does this look authentic or not, independent of whether it passes the sniff test, so to speak, yes, I think that type of technology is out there.

The Chairman. And it won't always be possible to disprove a video or audio by disproving the circumstances around it. In other words, if there were an audio of Dr.

Wenstrup purportedly on a phone discussing a bribe, Dr. Wenstrup wouldn't be able to say that I was in this place at this time and I couldn't have possibly been on the phone because the call could have taken place at any time. Or if there is a video of Val Demings, it won't always be possible for Val to show that she was somewhere else at the time.

Do you see the technology getting to the point where, in the absence of the ability to prove externally that the video or audio is fake, that the algorithms that produce the content will be so good that the best you will be able to do is a computer analysis that will give you a percentage? The likelihood this is a forgery is 75 percent, but you won't ever be able to get 100 percent.

Are we headed for that day where it just won't be possible to show that something we have seen or heard is illegitimate?

Mr. Doermann. So part of the MediFor program was exactly that, coming up with a quantitative scale of what manipulation or what deception is. I don't know if they have gotten there, I left partway through the program.

But, yes, I think there is going to be a point where we can throw absolutely everything that we have at this, at these types of techniques, and there is still some question about whether it is authentic or not.

There is no -- you know, in the case of the audio, you could do close analysis with tools and voice verification, all of those sorts of things. But just like a court of law, you are going to have one side saying one thing and you are going to have another side saying the other thing, and there are going to be cases where there is nothing definitive. I definitely believe that.

The Chairman. Do my colleagues have any further questions?

Well, on that optimistic note, we will conclude. And, once again, my profound

thanks for your testimony and your recommendations.

The committee is adjourned.

[Whereupon, at 11:15 a.m., the committee was adjourned.]