THE BROOKINGS INSTITUTION
Brookings Cafeteria:
Friday, August 16, 2019

PARTICIPANTS:

Fred Dews
Managing Editor, New Digital Products
The Brookings Institution

John Villasenor
Nonresident Senior Fellow, Governance Studies
Center for Technology Innovation
The Brookings Institution

(MUSIC)

DEWS: Welcome to the Brookings Cafeteria, the podcast about ideas and the experts who have them. I'm Fred Dews. Deepfakes are videos that make a person appear to say or do something they did not say or do, and they are coming to an election near you. With the 2020 election contests coming up, how can we guard ourselves against deepfakes and prevent them from changing the outcome of an election? To address this problem, I'm joined in the Brookings Podcast Network studio by John Villasenor, a nonresident senior fellow in Governance Studies in the Center for Technology Innovation at Brookings. He is also a professor of Electrical Engineering, Public Policy, Law, and Management at UCLA.

You can follow the Brookings Podcast Network on Twitter @policypodcasts to get information about and links to all of our shows, including Dollar and Sense, the Brookings trade podcast, The Current, and our events podcast. If you like this show, please go to Apple Podcasts and leave us a review. It helps others find it. And now, on with the interview.

John, welcome back to the Brookings Cafeteria.

VILLASENOR: Thank you very much.

DEWS: It's good to have you back in the studio. I briefly described in the intro, but can you offer more detail to explain what a deepfake is?

VILLASENOR: Right, so a deepfake is a video that has been generated with the aid of artificial intelligence methods, and it is a video where a person is portrayed saying and or doing something that they never in fact said or did. And the technology for doing this has become dramatically more advanced in recent years.

DEWS: Well, if it requires advanced technology, does that mean that they’re going to be rare or are we going to see more and more of them as time goes on?

VILLASENOR: It's a great question, and technology is certainly very advanced relative to where it was a few years ago, but it's not hard to get. In fact, anyone can do it. Anybody who has a computer and access to the Internet is in a position to produce deepfakes. So, the landscape has really changed very recently in that respect. The last election cycle, for example, that was not the case.

DEWS: Why do you think that they’re becoming more prevalent recently? Is it just because technology is easier to get?

VILLASENOR: The technology for producing deepfakes has become much more widely available than it was in the 2016 election cycle. There is a lot more public awareness that these things exist, and that of course means that the people who might make them are also more aware of them than they were in previous years.

DEWS: I’ll just point out saw kind of an interesting thing on Merriam Webster's Twitter feed. They tweeted at the end of July that the term “deepfake” is a word that they're watching for fast tracking into the official dictionary.

VILLASENOR: I think they should add it to the official dictionary. The usage has gotten ahead of the dictionary.

DEWS: So, what are some of the ways that you have seen deepfakes being used?

VILLASENOR: Well, they're used, for example, to portray politicians perhaps appearing to say something that they didn't say, and I should also say that they're not always a negative thing. So, for example, there's an art museum in Florida which is using deepfakes of Salvador Dali so that the visitors to the museum can interact with the deepfake and learn about the artist and his work. So, that's obviously not a negative use, but they can also be used in a negative way, of course, if they're constructed to damage someone's reputation.

DEWS: In June, you wrote a piece for the Tech Tank blog on the Brookings website, and I'll quote, “To influence an election, a deepfake doesn't need to convince everyone who sees it. It just needs to undermine the targeted candidate's credibility among enough voters to make a difference.” Can you expand on that point?

VILLASENOR: Yes. The point was that -- because many people look at these deepfakes, because some of the deepfakes today you see are very realistic, but some of them are clearly manipulated and really obvious as fake. But the point is that if you have a deepfake that's been distributed to, let's say millions of people, then it doesn't have to convince everybody, especially if it's a deepfake that is engineered to play off of negative biases and expectations that they may already have about a political candidate. And so, if you look at how that might propagate through the ecosystem, it could still be very damaging and very effective in a negative way, despite the fact that a cold, objective analysis would clearly reveal it to be a manipulated piece of media.

DEWS: I just recently did a podcast interview with Darrell West, the VP and Director of Governance Studies, and he kind of addressed that this particular issue, in terms of political polarization, that political partisans may not care if a video is quite obviously faked if it suits their particular partisan aims.

VILLASENOR: That's right, and not only might they not care, they might want to amplify a video that they know is fake for the reason that it does suit their particular aims, even though they know it's fake.

DEWS: Let's talk about some of the ways to address deepfakes. There's technology, there's legal approaches. Can you unpack some of those, please?

VILLASENOR: Right. So, there's really a multipronged set of approaches to address them. Unfortunately, none of them perfect. But on the technology side, there's quite a bit of work going on for deepfake detection. In other words, to develop algorithms where you can feed the algorithm the video and then the algorithm would look at the video and say, “hey, this media looks like it has been manipulated. It's probably not genuine.” So, there's a whole set of techniques there.

There's also legal measures. In other words, if a deepfake has been placed into the stream of social media, there are various legal frameworks that could be used, including potentially copyright law, the right of publicity, the torts of defamation, false light, and intentional infliction of emotional distress. So, there is a set of tools, but also with all of these tools, there are potential advantages but also potential drawbacks and things that act in tension that limit the effectiveness of these tools.

DEWS: You have a really interesting point in your piece, and it goes back to the Salvador Dali and the art museum example, that if we had some kind of a programmatic way to spot and eliminate deepfakes on say, Facebook, what about the deepfakes that are intended for beneficial reasons rather than malicious reasons?

VILLASENOR: Right, of course we wouldn't, if for example, an educator has made a deepfake that is providing an educational purpose and that in no one's mind is nefarious or malicious, then it would obviously serve no good purpose to have social media companies automatically identify and remove that deepfake from their systems. And I also will say that the fact that a deepfake might make the person it's portraying not particularly happy, doesn't necessarily mean as a matter of policy it should be removed. For example, parody, right? Parody is obviously a protected form of expression and so one can imagine that at some point, deepfakes will be used for parodies, and as long as they are not used in a deceptive manner, that's not something that we should force people to remove, even though it might be viewed in a negative light by some of the people depicted in the parody.

DEWS: It strikes me that any political campaign that created a deepfake video portraying their opponent in a negative light could just claim, “well, it's just a parody. Of course we don't really mean that they said this is what they said.”

VILLASENOR: And I think a bigger threat in the political election context is not necessarily that a campaign itself would hire people to create a deepfake, but that they would give sort of a wink and a nod to their supporters to produce and distribute these things. And so, I think that's the bigger risk. Or, that some of their supporters might do that anyway, of course without any involvement or direction from the campaign. The Law of Large Numbers really ensures that's going to happen if you have tens of millions of people who are supporting a particular political candidate, there's going to be some subset of them who think that creating and launching a deepfake targeting their opponent is a thing that they want to do.

DEWS: I want to stick on this question of legislation for a few more minutes. And by way of getting there, I just want to say I noticed that Senator Josh Hawley of Missouri, he just introduced legislation to force social media companies like Facebook and Twitter to change the way their tools function and also how users interact with them. But that strikes me as addressing the wrong kind of problem. But what would legislation look like to address deepfakes? Who would it be aimed at? Who would be sanctioned in creating deepfake videos?

VILLASENOR: Yeah, well first of all, I'm not a supporter of the legislation that you just mentioned. I don't think it's the role of Congress, for multiple reasons, to micromanage the user interfaces that social media companies present to their users. With respect to legislation, it's a complex issue with deepfakes, because the challenge is it's relatively easy to draft legislation that would help stop the proliferation of malicious deepfakes. But it's hard to do so while also not running afoul of the First Amendment, and by the way, I'm not saying that all deepfakes should be protected under the First Amendment. There are certainly lines that can be crossed, but it's hard to draft legislation with the requisite precision to sort of thread that needle, and so that is going to be a challenge. Let me also say, I'm always in favor of being cautious with respect to new legislation to make sure it actually solves a problem that isn't already solved with existing frameworks, and I think there's a long list of legal frameworks already on the books and we haven't seen evidence that they have failed in the context of deepfakes.

DEWS: Do you think social media companies are interested in trying to combat deepfakes on their own? And also, when I said social media companies, I used to think about Twitter and Facebook, but are there other social media platforms that we ought to be thinking about?

VILLASENOR: Well, YouTube is another example. But certainly, Twitter and Facebook -- I think they are. I think they're very interested and they're very much aware. I don't speak in any official capacity for them, but just I would imagine that given the attention on just misinformation, even if we go back before deepfakes. Certainly, misinformation was a top of mind issue in the 2016 election cycle in respect to social media, and so they're very sensitive to the problems with misinformation and deepfakes, or sort of the unfortunate next step in misinformation. I would expect that they're looking very carefully at how they address it. But again, the challenges -- it's a separate issue from the legislation issue -- but the policy issue, within a social media company, it's also not easy to come up with a policy that makes sense, because for example, we don't want to remove the Salvador Dali deepfake that was created by a Dali Museum for educational purposes.

DEWS: So, I don't think that deepfakes were used in the 2016 election. I don't think they were used in the 2018 midterm elections, and clearly they're going to be seen in the 2020 elections. Can you talk about what your expectations are for in what campaigns are we're going to see them in 2020? Just the president, or other elections, and what the prevalence of them might be?

VILLASENOR: You know, it's hard to tell in advance. I think clearly, the higher profile the election or the particular race that we're talking about, than the more people there are who are going to be engaged and just statistically, then the more risk there is that there will be people or supporters of one candidate or opponents of another who are going to actually engage in producing deepfakes. You can also imagine that it will be a tool in the toolbox of state actors, to the extent that state actors might seek to influence, in some way, the 2020 U.S. presidential election. It's not hard to see that that might happen as well.

DEWS: So, we've talked about addressing the problem of deepfakes through legislation. We've talked about what social media companies can do on their own. What can voters do to combat, if they want to, against deepfakes? How can we spot a deepfake?

VILLASENOR: Well, I think that's a great question. I think all of us, and I think I would say not just voters, because deepfakes can be a concern of course in areas outside elections as well. So, I think as consumers of digital media, which all of us of course are, I think it's important for us to, in some sense, unlearn what we've learned, you know, since we were all small, which is usually seeing is believing, right? If you see something with your eyes, then you know you can have high confidence that what you're seeing is actually a representation of what's real. Deepfakes scramble that understanding. They make us question that as well they should.

And so, as voters, if we see a video of a candidate engaged in saying something that just seems so uncharacteristic of that candidate that you think, “oh my gosh it is that person actually saying?” we should not necessarily jump to the conclusion that it's real video and do some diligence, which hopefully a quick internet search would provide, about whether there's a concern about whether that video has been manipulated or not.

DEWS: But do you worry about the technology getting so good that we might not believe the words that are coming out of the speaker's mouth, but we just won't be able to tell if like, that's really their mouth moving?

VILLASENOR: No, you're right. The technology, I mean sometimes when I speak to people about this, they look at examples and they say, “well, these are easy to spot as fake.” But remember that we've gone from a world four years ago, where there were essentially no deepfakes, at least in the broader public dialogue, to one where a quick internet search will surface dozens of them. And so, technology will continue to advance.

And you're right, it will be possible in the near future and today, even, if you have enough resources, to produce deepfakes that are just extraordinarily well produced. Many of them are not going to be in that category, but some of them will be. So, even if you know it's fake, it can still be powerful, right, even if you know something is fake. The visual imagery is very powerful and so I think it's a big concern.

DEWS: Well, John, I want to thank you for taking some time out of your schedule to talk to us today about deepfake videos and what we can do about it.

VILLASENOR: Well, thank you very much. It's a fascinating and important topic and I'm sure we'll be talking more about it.

DEWS: Yes, we will. You can learn more about John Villasenor and his research on our website or visit his webpage at UCLA.

The Brookings Cafeteria Podcast is the product of an amazing team of colleagues, starting with Audio Engineer Gaston Reboredo and Producer Chris McKenna. Bill Finan, Director of the Brookings Institution Press, does the book interviews and Lisette Baylor and Eric Abalahin provide design and web support. Our intern this summer is Betsy Broaddus. Finally, my thanks to Camilo Ramirez and Emily Horne for their guidance and support.

The Brookings Cafeteria is brought to you by the Brookings Podcast Network, which also produces Dollar and Sense, The Current, and our events podcasts. E-mail your questions and comments to me at bcp@brookings.edu. If you have a question for a scholar, include an audio file and I'll play it and then answer on the air. Follow us on Twitter @policypodcasts. You can listen to the Brookings Cafeteria at all the usual places. Visit us online at brookings.edu. Until next time, I'm Fred Dews.