After the Fact | Looks Can Be Deceiving: Deepfakes

Originally aired January 18, 2019

Total runtime: 00:19:42

TRANSCRIPT

[Music fades in.]

Hany Farid, professor, Dartmouth College: Suddenly there'll be the ability to claim that anything is fake. And how are we going to believe anything? And that, I think, is a real fundamental threat to our democracy. Because I don't know how we have a democracy without being able to believe what we see and hear.

Dan LeDuc, host: That's Hany Farid. He's a computer expert at Dartmouth College who has advised the government and news organizations about a new technology that could threaten journalism, national security, and even democracy itself—the "deepfake."

["After the Fact" theme music plays.]

Dan LeDuc: From The Pew Charitable Trusts, I'm Dan LeDuc and this is "After the Fact."

Democratic government—and smart policymaking—depends upon the ability to make decisions using a common set of facts. But in today's world, sorting fact from fiction is harder than ever.

Our data point for this episode is 57 percent. Fifty-seven percent of social media news consumers expect the news they see there to be largely inaccurate.

But there's inaccurate because of errors and mistakes made by journalists and others. And then there's whole-cloth deception.

I'm talking with Hany Farid, a man some have called the "Sherlock Holmes of digital misdeeds."

Hany Farid, welcome.

Hany Farid: Thanks. Thanks for having me.

Dan LeDuc: Let's start by helping people understand exactly what it is we are talking about. What is a deepfake?

Hany Farid: So we're all familiar with fake images, fake video, and fake audio as well. And we've all seen photoshopped images. We've all seen fake video on YouTube. And the deepfake is the general category of the automation of the creation of that, using advances in artificial intelligence and machine learning.

Dan LeDuc: And a lot of times we see fake videos, and we know they're fake. They're sort of intended to be entertaining or something. That's not what we're talking about here. We're talking about something much more insidious.

Hany Farid: Yeah, I think that's right. We're not talking about images or videos of sharks swimming down the street after a hurricane. We are not talking about funny videos of basketballs going 300 yards into the hoop.

We're talking about the whole-cloth creation that depicts people doing and saying things that they never did. And in some cases, that's done for humor—splicing in, for example, Nic Cage into "The Sound of Music." Very funny, very entertaining.

Dan LeDuc: Troubling to some, maybe.

[Laughter]

Hany Farid: Yeah, maybe disturbing, but funny. But then there are more nefarious purposes like, for example, the creation of presidents, candidates saying and doing things that they never did. And that is the real concern for us.

Dan LeDuc: We're talking about something that is really, really new. And I saw a quote from you that said, "It's always going to be easier to create a fake than detect a fake." So why is it easier to do this? Why can't we see it and detect it?

Hany Farid: Yeah, the nature of almost any authentication game, whether that's currency, art, documents, images, video, audio—

Dan LeDuc: Let me interrupt. I mean, we're talking like, counterfeiting. So that's a good analogy, maybe, for people to think about a little bit, right?

Hany Farid: Yeah. I think that's probably the best analogy. Because for years—in fact, for decades—we have been fighting this counterfeit problem. And it's a good analogy for a couple of reasons. One is that you have this two sides of the equation. You have the adversary and the defender. And also you have a monetary incentive to create the fake.

And also in the world we're living in now with deepfakes, there are monetary—and in fact, we are seeing state-sponsored attacks.

So think about, for example, back into the cyberspace. Spam, malware, viruses, ransomware. We get better at defending against them, but they're still there. I still get spam every single day. And that's because our adversary is constantly adapting. And that means we have to constantly adapt. And it's very much an arms race. It very much is constantly escalating. And you never really solve it. You sort of control it a little bit. You minimize the damage. But it's always there. There's always that threat there. And I think that's probably the game we're going to play with this type of content as well.

And it's particularly nefarious now, and it's particularly problematic now, because the content is being created by computers. So we've taken, in many ways, the human out of the loop. So the time, and the expertise, and the expense to create the fake is largely gone now. And the computer, the artificial intelligence system, has just taken over. And so that makes the adversary's job even easier now because it's not like they have to be particularly skilled to create this fake content.

Dan LeDuc: Well, let's spend a moment and describe for folks what it is we're talking about. People only need several hours of someone talking. What's the raw material that someone needs to create one of these?

Hany Farid: Yeah, so here's the really simple—let's do a couple of versions of this. So imagine you wanted to create a video of a movie star or a politician inserted into a video that they weren't originally in. So what you need, of course, is the original video. That's easy to get. You just take whatever video you want. Whether you want that to be "The Sound of Music," whether you want it to be a press conference, whatever it is, that's your target video.

And then, depending on the techniques that you're going to use, you need anywhere from a few hundred to a few thousand images of somebody. Now for famous people, that's actually relatively easy to get. You just scrape YouTube. You scrape the web. You can find hundreds, thousands, tens of thousands of images of them. And essentially all you really do is you feed an artificial intelligence—or more specifically, a machine learning algorithm—all of that data. And you instruct it to replace the face of the person in the original video with this person's face.

And what it does is it essentially searches the space of all the appearances of the person that you've given them. And it knows what the person looks like in the video. And it either splices one in, or it synthesizes one if it doesn't have it. And then it splices it in. And it does that frame, after frame, after frame in the video. And that's how you create the fake video.

And it actually does it fairly well. They're not perfect. But this technology—every three to six months, we see rapid advances in the realism and the sophistication of the types of fake videos that it can create.

Dan LeDuc: Wow. Every three to six months. Which means if you and I are talking a year from now, and nothing is done to sort of come up with alternatives and fixes, these things could look seamless.

Hany Farid: Here's what I can tell you. If we were having this conversation a year ago, I would say that this was barely on our radar screen. So here we are a year later, and it is of great concern. And—there's a big if—but if that technology continues to develop at the pace that it has, and all evidence is that it will, I agree with you that a year from now we are going to have the ability to create very sophisticated, very compelling, and difficult to detect fakes. And we will have democratized access to the technology that does that.

So suddenly the threat is actually dual now. Because now suddenly there'll be the ability to claim that anything is fake. And how are we going to believe anything? And that, I think, is a real fundamental threat to our democracy. Because I don't know how we have a democracy without being able to believe what we see and hear.

Dan LeDuc: And we're also at a time when the public is increasingly skeptical of what they see online. So we almost seem to—the third factor is almost like this ripe audience for this sort of thing as well.

Hany Farid: We have a particularly polarized public, both here in the U.S. and abroad. Part of that is just the nature of our politics today. But part of that is also the echo chamber that is social media, that we are now seeing constantly reinforcing media because of the way content is promoted to us.

And that's, in some ways, the perfect storm: We have the ability to create the content, we have the ability to distribute the content, and we have a willing public to consume it. And if any one of those was disrupted, it would still be a threat but not nearly as large of a threat. So I think the fact that all three of these are lined up like this is a particularly dangerous time for us.

Dan LeDuc: You know, our data point for this episode is that 57 percent of social media news consumers say they actually expect the news they see on the social media to be largely inaccurate. That really speaks to the mindset of much of the public. But I do wonder if they're really thinking about, is the news they're seeing purely a deepfake?

Hany Farid: Yeah, I wonder what people think when they say "inaccurate." Because what we know is that it's not just inaccurate. In some cases, it's just outright false. It's just completely made up, and sometimes intentionally so. I find that number, by the way, incredibly disturbing. And I find it particularly disturbing when you look at another number, which says that now more than half of Americans get the majority of their news from social media.

So put those two numbers together and I think that is something that we should be very, very concerned about.

Dan LeDuc: And let's emphasize that this is not some scientists at places like Dartmouth and Berkeley and elsewhere who are just sort of talking about hypothetical threats. The broader potential threat at the sort of world political level has led the government to take this seriously. You're working with the Defense Advanced Research Project. That's sort of the Pentagon's think tank, right?

Hany Farid: Yeah, that's right. So DARPA, about four years ago, started getting serious about developing technology to detect fake content. And I will say that this was before the explosion in artificial intelligence and machine learning. This was before the word "deepfake" was even in our vernacular. But we were already concerned about the ability to create fake content now, what we considered the good old-fashioned way, with things like Photoshop.

We are particularly concerned about the national security issues. We're concerned about how this can be used to disrupt national elections. We are worried about how media outlets will be able to vet information. We are worried about the threat to movie stars, actors, politicians, average person, and your ability to create harmful fake content with them in it. And across the board, we are very concerned about the proliferation of this type of content and what it means to us as a society.

And you said this—and I think it's important to emphasize that—this is not abstract. We are seeing real consequences to this. We have seen violence and death in Myanmar, and Sri Lanka, in India, in the Philippines that have been around fake news and fake content. And so these aren't abstract notions anymore. We have seen election tampering, both here and abroad, in terms of interference from outside parties.

So these are not just academics sitting around saying, "Hey, this is a possible threat." We are seeing real-world implications of this. And I think it's the tip of the iceberg.

Dan LeDuc: What can we do, and what's being done?

Hany Farid: One of the things that we are working on is developing technology that can, for example, determine that a video of President Trump, or Prime Minister May, or Chancellor Merkel is in fact them.

And I'll just give you some sort of broad strokes on how we do that. So one of the things that we've discovered watching hundreds of hours of our world leaders is that they have particular tics, if you will, or patterns in how they speak.

And so one of the things we've been doing is building what we call a "soft biometric" that looks for these patterns in how we talk, how we raise our eyebrows slightly when we say a word, how we tilt our head slightly when we, for example, are contemplating something or are upset about something, or how we, for example, smile and how our eyes move and our lips move and our head moves when we do that. And what we are finding are these very strong correlations and, more importantly, that those correlations are disrupted in the creation of a fake. Because you're sort of creating this chimera of a person. You're putting words into their mouth. And you're changing facial expressions that are not consistent.

And even though we as humans may not necessarily detect that—but the computer can measure these things extremely accurately and then look for deviations. And so that's one of the things particularly with respect to world leaders that we are very concerned about because of the influence—the undue influence—that that can have on everything from national security to democratic elections.

Dan LeDuc: And that gets at authenticating a video like this. But as you were talking about earlier, if someone decides to create a video of a world leader who says, "We're under attack, and we're launching our nuclear weapons"—that's going to go viral around the world in seconds.

Hany Farid: A nanosecond.

Dan LeDuc: Yeah. I mean, millions of people will see that in the time it takes me to finish the sentence. And so how do we address that speed—I mean, what do we do?

Hany Farid: This is exactly the right question. So for example, if somebody sends me a video, I may be able to analyze it hours, days, weeks after its release and come to some conclusion. OK, fine. But what do you do when you've got five seconds to make a decision on the authenticity of video? And while that doesn't happen often, when it does happen, it's going to be serious.

Dan LeDuc: The stakes are very high in those moments, yes.

Hany Farid: So now we have an interesting problem. But let me make the problem a little bit worse for you. Today on YouTube, there are between 400 and 500 hours of video uploaded every minute. So now you've got to deal with that insane amount of data being uploaded on a daily basis and be able to run technology, if you want to do it on upload, essentially instantaneously. And that is a huge ask.

So I don't have a good answer to that. I don't think that any technology that we develop in the next few years will be able to work at that scale. So the best we can hope for is to put this technology in the hands of the National Security Agency, the CIA, the FBI, the law enforcement agencies that are, in the end, going to have to make these decisions, so that when these videos are released, there is somebody in power who can have the tools that they need to make an assessment.

And this is essentially that same type of questions we've had with physical weapons. How do we distribute these things in a way that protects us, protects our allies, but doesn't arm our adversaries? And I think it's a very analogous position. And I don't have great answers to this. It's constantly changing. The landscape is constantly changing. So an institution we may trust today, we may not trust tomorrow.

Dan LeDuc: Let's go back to technology for one last moment, which is—can you talk us through a couple of things that are in the works to help address this in the sense of how videos can be created and authenticated even when they're being created?

Hany Farid: So I think there's a couple of ways to tackle this problem. So let's say there's two basic ways. So one is the way that I've been talking about, which is: We are given a video. And we have developed technology to detect that, whether it's looking for patterns in the way somebody speaks or maybe we are finding artifacts in the video that are results of being generated by a computer.

Now, there's a fundamentally different approach to this, which is that, when there are mission-critical videos—for example, a political speech or a video of police misconduct, or human rights violations, or natural disasters—that those videos are recorded in what is called a controlled capture environment. And so the idea there is instead of using sort of the standard camera video app that you use on your mobile device, that you use a special type of software that sits on your iPhone or your Android or whatever your device is.

And what it does is at the point of record, and exactly at the point of record, it cryptographically signs the content, puts that signature on the blockchain, which is essentially a distributed ledger that makes it very difficult, if not impossible, to manipulate that signature. And then that allows us later on down the line to

authenticate that video by comparing the signature to the signature that was put on the blockchain at the time of recording.

It's a very nice technology. It puts the burden not on the person receiving the video, but on the person creating the video. So in the high-stakes situation particularly, this is a very nice technology. And that technology exists today. And I'm a strong advocate of deploying this type of technology because it's something that we can do today. It works at scale. It's highly reliable. And I think it will deal with some of the concerns, that when you have this control capture video we can say something fairly definitive about it. And then we'll deal with the other situations as we have to.

And so I'm really encouraged by that. I will say that on the other side, the more forensic analysis side, people like me are outgunned. There is probably 100 to 1,000 to 1 more people developing the technology to manipulate content than there is to detect.

Dan LeDuc: Hmm.

Hany Farid: I'd like to see more people doing what I do. I think that will help. But I think also we're going to have to come to grips with how we deploy that technology, how we release that technology, who has control of that technology.

Ideally what I would do is I would tell you there's a technological solution. That there's a button you can push that will authenticate content for you. There's not, and won't be for a while. So now we revert back to the human component.

Dan LeDuc: So maybe we can offer some advice to just the average person who sits down and looks at their laptop or their phone, which they do countless times a day now. Knowing that this is a possibility out there, what can the average person start doing?

Hany Farid: I think that there's a couple things that we as the consumer, as the public, can do as we consume content online. Number one is slow down. The fact is the internet moves very, very fast. We move very, very fast on the internet. We are blowing through things, largely reading headlines, retweeting, liking, sharing, without really digging in. And we have to slow down. We have to slow down in how we consume content.

We also have to become more critical without becoming cynical. We also have to understand that, particularly on social media, the content that is being fed to us is going through a filter. It is being catered to you. And we should understand that. And that's really important. So understand that the content being provided to you on social media is being provided to you to keep you on that platform for as long as possible. And so we have to get out of that echo chamber. We have to understand that there are diverse

views out there, even if we don't agree with them. And I think become more critical and thoughtful.

I also think that there is very, very good fact-checking out there. There's Snopes.com. There's PolitiFact. There are many, many websites out there that try and do a fairly good job of quickly dealing with rumors and false stories. And it just—it doesn't take that long to look those up before we are sharing and retweeting and liking. And understand that when you do that, you are part of the problem. You are part of the fake news problem, you are part of the virus, you are spreading the virus. And we need to, as individuals, take responsibility for our part in that.

[Music fades in.]

Dan LeDuc: Hany Farid, thank you so much. And let's tell our listeners that, yes, you, in fact, were sitting in front of a microphone at Berkeley. And I was sitting at a microphone in Washington. And this conversation was actually what they just heard.

Hany Farid: I will attest to that.

Dan LeDuc: Thanks.

Hany Farid: Thank you very much.

Dan LeDuc: If you want to learn more about how this works and see an example of a deepfake, visit our website at pewtrusts.org/afterthefact.

There you can watch a video that claimed to show a commercial airliner doing a 360-degree flip before landing. You won't believe your eyes—and you shouldn't.

You can also find previous episodes related to this conversation, including our chat with Arati Prabhakar, former director of the Defense Advanced Research Projects Agency, about the power—and the peril—of new technologies.

Thanks for listening. For The Pew Charitable Trusts, I'm Dan LeDuc and this is "After the Fact."

[Closing "After the Fact" theme music plays.]