SEI Cyber Talk (Episode 13)

Deepfakes—What Can Really Be Done Today?
by Rotem Guttman and Zach Kurtz

Rotem Guttman: Hi, this is Rotem Guttman, and I'm going to be speaking with a fellow researcher, Zach Kurtz, about deep fakes, what's real and what's fake. So Zach, to jump right in, what are deep fakes?

Zach Kurtz: All right. Well, a deep fake is any kind of-- typically photos, but it really could be any kind of essays or any media-- that has--

Rotem Guttman: Or videos.

Zach Kurtz: Videos or photos, right, that has some fake content in it, and it's not just any fake content. There's all kinds of ways to just throw some weird pixels or content into a bit of media, but a deep fake uses a deep neural network which is a kind of machine learning thing where they can generate fake content and it's really taken off in the last few years. And so the other aspect of it is it's not just how you make it, but how you use it. So you could imagine using some of this modern machine learning stuff to do red eye removal or do other special effects on images that serve a simple artistic purpose or help to make the image more truthful to the original goal, but with a deep fake, it's being used to trick or deceive or even to humor people, and that's part of the important part of what makes up a deep fake.

Rotem Guttman: How are these deep fakes created? How are they made that's different from just traditional Photoshop or aftereffects?

Zach Kurtz: Basically they need to come from deep learning somehow. So a deep learning model is a statistical or machine learning model that has hundreds, even millions, of parameters, and you trained it on a whole bunch of real content and real essays or real videos. If you're doing video, it's going to be trained on images or videos, and it kind of learns-- this model learns how to make new content, after lots of training, if you train it in a particular way. Now, there's many different kinds of deep fakes, and exactly how you make each one varies widely depending on exactly what you're trying to do.

Rotem Guttman: You mentioned that there are different kinds of deep fakes. What are the different kinds?

Zach Kurtz: Sure. I mean, most broadly, you can fake anything. You can fake a sunset or fake clouds moving across the sky. Maybe one of the first things people might recognize as deep fakes was when Google put out Deep Dream, and they would make these psychedelic images that sort of synthesized content from many images that were used to train the model. But when you get to more-- what people have really been paying attention to lately is these deep fakes of people talking, especially famous world leaders who are saying things they wouldn't really say.

So how do they make those? Well, there's several different kinds, and they mostly focus on the face and replacing bits of the face or all the face, and so what exactly are they replacing? And the smallest kind of replacement would be you just-- you take existing footage of a leader or whatever the target is talking and you just replace the section right around their mouth and replace it with-- make the lips look like they're matching up with some other audio, and that's called lip sync. The target ends up lip-syncing to whatever you want them to say. And so maybe the most famous example of that is when Jordan Peele impersonated Obama's voice so accurately, they were able to do a lip sync and made Obama look like he was saying was pretty ridiculous things, which was funny. So that's lip sync.

If you want to make someone look like they're doing something ridiculous, make a scene of somebody doing something ridiculous or saying something ridiculous, and then just remove the head of that person and put the puppet, or the person you want to have look ridiculous, over top. And so you replace the head. That's called a face swap.

Rotem Guttman: But there you need your actor doing something ridiculous to be a body double, essentially, for the person who's face you're pasting on.

Zach Kurtz: That's right. So you need to have some certain plausibility that that actually looks like it could be that person's body. And you can do that in non-deep ways. There's shallow fakes that do face swap, and then you can also use deep fake.

But I think were deep fakes really get scary and interesting is with what they call puppet master, and in puppet master, the puppet is the person you're trying to make look silly, and the master is the actor or the raw footage showing the scene of somebody doing something silly. And in that case, it's a little bit like face swap but it's a bit more powerful, where instead of just totally replacing the head, you regenerate the puppet, target person, in every frame of the video to be doing what the master's doing, and matching everything-- their entire facial expression-- but in the identity of the puppet instead of the master, and that you can get the puppet to be saying exactly what the master was saying but in the puppet's voice. So that actually raises the point-- I mean, there's a whole bunch of techniques for faking imagery. There's a comparable set of techniques for faking voice and putting them together in a synchronous way is a bit of a challenge, but people are starting to do that.

Rotem Guttman: But at the moment at least, those are two separate systems, essentially-- faking the face and faking the voice.

Zach Kurtz: Well, largely. Now, a very recent paper has looked at piecing together phonemes and what they call visemes, or bits and pieces of how your face is shaped and how it matches up with the sounds you make, and putting those all in one big joint model and try to do it at one time. But that's really, really cutting edge.

Rotem Guttman: What's really the difficulty level right now for each one of these? If I have some really good quality footage of somebody I want to impersonate, I happen to be a really skilled voice actor, and I can impersonate their voice convincingly, what's it take for me to lip sync? And then at the other far end, what's it take for me to fully replace somebody's actions and voice and everything and create a brand-new scene to make a target look like they were involved in it?

Zach Kurtz: Starting with the both basic, where you're just doing a lip sync, replacing a bit of their mouth and you impersonate the voice-- I mean, voice impersonation could be hard, but if you're good at it and get a good recording-- you hear in the news that it's super easy now. I'd say that's a little bit overhyped. Maybe if you have a PhD student working in computer vision, they might be able to hack it together in a couple of days to get what you're going for.

Now if you want a full-out deep fake where the whole thing is imaginary, that might be doable, but it's a major project.

Rotem Guttman: That being said, what makes it a major project? What are the stumbling blocks nowadays that are preventing this from just being a turnkey solution where you can make anybody do or say anything?

Zach Kurtz: There's just a lot of moving parts. You've got to synchronize the voice with the image of the person moving, and if you're going to-- so for example, if you were to have a scene of a person-- say it's a puppet master kind of scenario, where you've got a master acting out a scene, and you want to replace that master with a puppet, if you take that master out of the video and put the puppet over them, maybe things won't line up quite right, or each time you generate a new frame of the puppet, the frame is just off a tiny bit, and so when you make a video, it's jittery the whole way, so you have to smooth that out somehow. So there's a whole bunch of little technical things like that that aren't entirely-- they're not completely worked out.

Rotem Guttman: All of the use cases I've seen so far are, at least for video, doing this on a frame-by-frame basis and then doing that smoothing, as you mentioned-- some of them. Have you seen anything where they're actually trying to generate cohesive video directly without essentially making a motion JPEG and then converting that?

Zach Kurtz: I think so, but I can't really speak to that yet.

Rotem Guttman: Okay. That being said, what do you see the near future and the far future of these technologies? What do you think is going to be capable in one year, five years, fifteen years?

Zach Kurtz: Good question. It's very hard to predict. I think we're going to see steady advances and within a few years I would expect there would be an app where you could pretty much-- let's see. So you would probably be able to act out a scene and put your best friend in your place and make it look like they were doing what you're doing. I think that's coming soon. It's not quite there yet.

Rotem Guttman: So we're not sure how far out, but not 15 years?

Zach Kurtz: Well, one key question is what kind of dynamism are you going for in your scene. It's a lot easier just to-- like look at me right now. I'm standing here very close. My lighting is constant. I'm not moving my head around that much. If you've got a simple video like me just talking to your face, that's a lot easier to deep fake than if you've got someone running through the forest and swinging and a machete and the machete goes across their face and interrupts the field of view and all that. That's a lot harder. So there's always going to be a frontier on deep fakes. Even 20 years from now, there's going to be some effects that may be very hard to do.

Rotem Guttman: Oh yeah. This video alone being out there makes us both great candidates for faking. So if you see either of us doing anything strange, who knows if it was created or not. And that brings up another question. What do we do as we move into this world where content can be created that never existed? How can we actually tell-- beyond just being critical of everything out there and not trusting everything, how do we tell what we can trust and what we can't trust?

Zach Kurtz: Yeah, that's a good question. It's actually a very old question that goes well back beyond video. I mean, if you think back to the very beginning of humanity when people first learned to talk, the first thing we could do is tell a lie. And then we learned to write and you've got all kinds of-- I mean, "fake news" is a recent term, but people have always been writing falsehoods for quite a long time. So how did we trust print media? Well, we went with standard news sources. For a while that was somewhat limited by only large, well-funded organizations could print, and so it wasn't that many different sources to choose from. But as we get into the modern age, for digital content, you have to think about what you're reading and who you're getting it from, and the same applies to video.

I guess for about a century, as photography came out, photographs were kind of special in that they served both as a way of communicating information and as a means of verification. You take a photo of a scene and that's evidence for a crime scene. But in the age of deep fakes, that'll change and it'll be more like who took the photo and can they verify they took it and all that. And so that's just part of the reality we'll have to work with.

Rotem Guttman: So it seems like we're going to have to adjust our expectations and something that we're all going to have to deal with as this progresses into the future. Is there anything else

you want to mention about deep fakes before we go? What's your favorite deep fake that you've seen so far?

Zach Kurtz: Oh, it's a great question. There's so many good ones.

Rotem Guttman: Fine, top three.

Zach Kurtz: Well, I mean, so if we're talking about-- many of them are very political. You have to watch out. There's a lot of mish-mashes where they take two world leaders and blend them together to look like they're one person and it's very confusing; you don't know what you're looking at. I find that kind of deep fake effect to be really entertaining. Of course the Obama and Jordan Peele deep fake was just downright funny. And I don't know what I would take for my third.

Rotem Guttman: My favorite one is I was looking at one use case where they had trained their neural network in order to take-- instead of facial features, they had extracted edges of various images and then generated other images from those edges. So they were basically taking the outline of a cat and then generating pictures of cats, or the outline of a handbag and generating pictures of handbags. And so what was nice is you could draw a really crude drawing yourself, a line drawing, of a cat or a handbag, and then that neural net would try and interpret that as, "Here's what the picture of that would look like," which with handbags was great. Apparently I can design really good handbags, because I'd draw a crude buckle and it would put a buckle there, and if I could draw a crude cat, it would make some sort of abomination that was terrifying, out of my nightmares. But all in all, I think it's a very interesting technology to play with and I'm eager to see where it ends up.

Zach Kurtz: Me too.

Rotem Guttman: So if you'd like more information on the work that we do, check out the link in the decision. This has been Rotem Guttman and Zach Kurtz. Thank you for watching.

If you enjoyed that video and you'd like to hear more from our researchers, hit the Subscribe button. We're constantly adding new conversations and insights with researchers from all across our different directorates.

Related Resources

Jordan Peele: https://www.youtube.com/watch?v=cQ54GDm1eL0

Various types of deepfake (face swap, lip sync, puppet master): http://openaccess.thecvf.com/content_CVPRW_2019/papers/Media%20Forensics/Agarwal_Protecting_World_Leaders_Against_Deep_Fakes_CVPRW_2019_paper.pdf

Related work at CMU: https://www.cs.cmu.edu/~aayushb/Recycle-GAN/

AI at the SEI: https://resources.sei.cmu.edu/asset_files/FactSheet/2019_010_001_539538.pdf

General Tutorial: https://www.alanzucconi.com/2018/03/14/introduction-to-deepfakes/