Center for Strategic and International Studies

TRANSCRIPT

The Truth of The Matter
“Distrust of Everything: Misinformation and AI”

DATE

Tuesday, July 18, 2023

Featuring

Tiffany Hsu

Technology Reporter, The New York Times

CSIS EXPERTS

H. Andrew Schwartz

Chief Communications Officer, CSIS

Transcript By
Rev Transcript
www.rev.com

CSIS | CENTER FOR STRATEGIC & INTERNATIONAL STUDIES

H. Andrew Schwartz: I'm Andrew Schwartz, and you're listening to The Truth of the Matter, a podcast by CSIS, where we break down the top policy issues of the day and talk with the people that can help us best understand what's really going on to get to the truth of the matter about misinformation and artificial intelligence.

We have a real treat today because we have Tiffany Hsu, who is a reporter on the technology team at the New York Times. She covers misinformation and disinformation after spending several years on the media desk of the Times. Previously she wrote about the California economy, which is pretty huge for the LA Times. Tiffany, thanks for joining us today.

Tiffany Hsu: Thanks for having me.

Mr. Schwartz: So, I want to start with an overarching question. How do you think AI, which is changing everything these days or seems to be, how do you think AI is changing the spread of misinformation?

Ms. Hsu: To me, it's all about scale and realism. Previously, a lot of these disinformation campaigns that you would see out there, these coordinated attacks had tells, right, often they were in languages or they originated in languages that aren't English. So, you'd see little grammar mistakes, little spelling mistakes, just little clues that kind of raise the hair on your arm a little bit. AI is advancing so fast that that's quickly not going to be a problem. A lot of these chat bots are able to execute translations almost in real time and also at scale. Which leads us to the second point. Previously, you had a lot of information being spread manually. You know, had people who were typing out these false narratives and conspiracy theories and posting them on Twitter and Facebook, but they were limited because they were people. AI can churn that out at great volume. These chatbots can offer variations that make it hard to tell that it's something that's written by a machine instead of a person. We've done stories on the flaws in a lot of the detection technology that's out there. It's very difficult to say for certain whether something is produced by a human or produced by a computer.

Mr. Schwartz: Even the services that you've reported on that are out there to detect artificially generated misinformation, they don't get it right all the time. Do they?

Ms. Hsu: Absolutely not. Stuart Thompson and I did a look into detection tools used for AI imagery specifically, but the same thing applies for text and for audio and especially for video. I mean, the technology is there. Sometimes it's effective, but often it's not. I mean, one of the key examples in our story was that we took an AI generated image of basically a monster, a Neanderthal that's like 20 feet tall, standing next to two regular sized

humans, and in our first pass, none of the detection technologies could tell that this was AI generated. Even though as a human you're looking at and you're like, this thing obviously is not real.

Mr. Schwartz: That's incredible. I, and I guess the overarching thing here is that what does this say for the future of public trust in things like elections, things like that pertain to national security, everyday things. What does this do?

Ms. Hsu: I mean to say that experts in these fields are running scared is not an understatement. I mean, people are petrified. We've got 2024, the presidential election coming up. You've already seen the use of AI. The RNC Republican National Committee used it in really one of their first attack ads pretty much right after Joe Biden announced his reelection campaign. In that ad, the AI was disclosed, but not especially clearly. It's a tiny type in a corner of the video that they release showing scenes of destruction and drug-induced crises. Ron DeSantis, who's a leading Republican candidate, also used AI imagery. He released images that seemed to show Trump being buddy buddy with Anthony Fauci. A lot of that is already confusing people. We see this phenomenon that's pretty striking on social media a lot where people will post images that the original poster will say is created by AI. So, they'll disclose it off the bat. I used Mid Journey or Dolly or stable diffusion to make this image. It is not real. And in the comments pretty quickly you see people saying, wow, this is so cool. What camera did you use? What aperture did you use? So, the fact is that a lot of imagery and a lot of text and a lot of audio and a lot of video is already being made by machine, and people are believing that it's real even when you tell them that it isn't.

Mr. Schwartz: I mean, that's incredible. And I guess what that also means is that people now and into the future are going to have a hard time knowing what to believe and not to believe.

Ms. Hsu: That's already the case. One of the consequences of this situation is what experts call the Liar's dividend. They worry about people manipulating the lack of trust in order to advance their own end. So, the primary example that gets thrown around is that a politician might have an incriminating photo release. Let's say a politician is caught like doing drugs somewhere. All the politician has to do in the age of AI is say, that's a fake photo. AI made that photo. That's not actually me. And who's to say that he or she is wrong? Because the detection technology is not up to speed. Because AI technology is so readily accessible and so easy to use, it's really up for a voter to decide whether or not the politician is lying or not.

Mr. Schwartz: What role do you think policy needs to play in mitigating the spread of misinformation as AI continues to emerge and morph?

Ms. Hsu: We're doing a lot of reporting on that exact topic. There are efforts to implement policy specifically around AI use in elections and also how to manage deepfake pornography, for example, which is one of the primary uses of deepfake technology at this point, or malicious uses of deep fake technology. I should say, but a lot of the policy proposals that are in play are pretty toothless. If you talk to experts in AI, many of them will say that politicians understand that AI is likely to be a threat, but they don't fully understand how the technology works. So you have several issues here, right? One is that the lack of understanding means that a lot of these policies are overly broad. Another problem is that let's say you get a targeted policy put in place, it's law, these are almost impossible to enforce because a lot of the creators of AI content are anonymous. And then there's just the additional issue of whether or not AI policy is ever going to catch up to the technology. This is the same issue with detection tech because the AI technology at its core is advancing so rapidly detection tech can't keep up, policy can't keep up any policy that you put in place because laws take so long to establish is going to be several years behind where the current technology is. So, what I've actually heard from a lot of people is that it's not going to be laws that reign this in, or at least it's not going to be laws written by legislators. It's going to be lawsuits; it's going to be the court system. AI creators, when they can be identified are going to get dragged into court and accused of defaming real people, for example, and that's going to create the deterrent.

Mr. Schwartz: So, I want to talk about the courts in a second, but just to stick with the policy and regulation issue, I've heard Kara Swisher, the eminent tech journalist and podcaster say no one has in Congress or an administration has ever regulated the internet to this day. So, if that's what's going to happen with AI, we could expect regulations to come in around 2060 or so. They do seem though, to be taking AI much more seriously than they've ever taken some of the regulatory issues with the internet. Charles Schumer just released his plan here at CSIS a few weeks ago. What do you think of the efforts so far?

Ms. Hsu: It's very clear that Congress and also local legislators are worried about this, and they should be. Sam Altman of Open AI has testified before Congress, Yvette Clark, Amy Klobuchar, a lot of members of Congress have spoken out about their concerns around this technology, and they're proposing legislation trying to push it through. There are countless organizations popping up around DC that are debating the various facets of AI and how to manage it legally. I don't think the problem is that lawmakers aren't aware that this could become a threat. They are aware, they know the problems that they don't really know how to go about getting it in check, and frankly, no one really knows. It's just, it's advancing too fast.

Mr. Schwartz: Sounds like a really perplexing set of issues that Washington's going to have to figure out a way to deal with. You mentioned the courts and one of your recent articles is about disinformation and fallout from a recent judge order involving some cases in Louisiana and Missouri. Tell me about that and what that means.

Ms. Hsu: Yeah, so that was Steve Lee Meyer's story, if I remember correctly. And I'm not sure actually how much that had to do with AI. I think that was more just broadly on misinformation and how much the government is able to communicate with social platforms like Facebook and Twitter about reigning that in, about taking down or flagging problematic posts. The way this could intersect with AI is that if the government can't say, Hey, that artificially generated photo of an explosion at the Pentagon is potentially dangerous, if the government can't say that, then the social platforms aren't necessarily incentivized to take it down. They can offer any number of arguments. This is experimental tech. This is an exercise of free speech to whoever posted that image is trying to see what kind of reaction they could get. That's all valid. We don't have to do anything, right? I mean, that image in particular is kind of a prime example. That image moved markets for a short amount of time. You could argue that those sorts of things become national security threats. And so the concern is that the Louisiana judges' findings could really cramp the government's ability to step in when the social platforms won't.

Mr. Schwartz: And misinformation researchers are really worried about this, aren't they?

Ms. Hsu: Oh, absolutely. If there are no checks against misinformation, you start to see a degradation in societal trust. People start thinking, why am I going through all the effort to discern what is true and what is not? I'm just going to distrust everything. You're just going to create a society of skeptics. And I mean, we're living in a world where media literacy is already very, very frail. There are few schools that teach it to any acceptable degree. And so, what's happening is that the generation of people who use social media a lot or really get content from any source are just learning that they can't trust the content that they see because they don't know how to discern what's real and what's not. And AI is really just going to complicate that.

Mr. Schwartz: So, speaking of national security, what do you think are the biggest worries, potential misuse of AI and foreign affairs? Things like Russia's invasion of Ukraine.

Ms. Hsu: So very early on in that invasion, there was a deep fake of Zelensky that started making the rounds. And off the bat, it was pretty clear that it was a deep fake. It was relatively crude. Zelensky, if I remember correctly, almost

immediately disavow it. But a lot of the researchers that I've talked to have said that's the beginning. That is a signal of what could become in the, let's call it a year and change. Since that happened, the technology used to make that video has improved exponentially. If you look at image generators like Mid Journey, for example, when we first started reporting on that earlier this year, six months ago, Mid Journey was having problems generating hands realistically. So, you saw a lot of humanoid images with really fat hands with seven fingers on them. This was a tell that AI was involved. I mean the, it's funny, we wrote a story about a mayoral candidate in Toronto that released campaign materials that featured a woman who is resting her head on one hand, she has another arm crossed, and then there's a third arm just waving around in a different colored shirt. But it's very easy to avoid things like that. Now, Mid Journey Five has essentially erased the hand problem. So, the concern is that if a Zelensky deep fake were to surface now, even if Zelensky disavows at this time, it could be realistic enough that people could say that he's lying.

Mr. Schwartz: And what about people mimicking world leaders like President Biden or Russia's leader Putin?

Ms. Hsu: There have been unconvincing deep fakes of Biden on TikTok for a while now, and those are using AI filters, very crude applications of AI. But on Twitch, there was, I don't know if it's still running, it might be. There was a 24/7 livestream of Biden debating Trump that was very realistic. This was down to the audio. So, it was deep fake video accompanied by audio, and it was just running in real time all day, all night. For several days. It was very clearly marked as being AI. But if you were glancing at this thing, or a friend sent you this thing and you didn't really take any time with it, you could be forgiven for thinking that it was real. And I mean, this livestream had these Biden and Trump avatars swearing, saying really inappropriate things. And all someone has to do is take a screenshot or screen recording of that, omit the part where it says that it's AI and spread it. And people would think it's real, because this is a big problem with misinformation is that it doesn't exist in a sanitized vacuum, right? Misinformation is often copied and pasted. It's often cropped. It's often screenshotted. It's manipulated away from its original form. So, it's incredibly hard to track.

Mr. Schwartz: And the issue of marking videos that say, this was made with AI, for instance, a campaign video. It's so little and small when we're looking at it, especially if we're looking at it on our phone. How is anybody to really pick up that it was generated by AI, even the most discerning people?

Ms. Hsu: I mean, that's a concern. I should say, that there are many valid uses of AI, very valuable uses of ai, and there are many people who are using AI to create really great and thought-provoking art. And there are also people who are using AI to make a point about misinformation and trust where

they're being very conscientious about labeling. But there's been research done that suggests that oftentimes people don't pay attention to flags. They don't pay attention to labels, especially if they're on a social media platform and they're scrolling really quickly, and they don't have much time and they're not really paying attention. And so, then you have the question of, well, okay, if labeling and flagging isn't effective in a real world setting, what do you do? And that's where you get people who suggest like watermarking images, having some sort of content provenance marker on an image from the time it's created. But there are issues with that as well.

Mr. Schwartz: What about voice? Does voice help?

Ms. Hsu: Voice is one of the most problematic forms of AI at this point because it is so easy to do. I mean, there are apps out there that let you turn a recording of your voice into Morgan Freeman saying the same thing. Many of those apps are free and voice is really one of the hardest expressions of AI to detect as AI.

Mr. Schwartz: Sounds like we have a lot to think about. Tiffany, as you're continuing to report on these issues, what are some of the things that we should be looking for coming at us in the future?

Ms. Hsu: We're going to be doing a lot of reporting on 2024. Election administrators are very concerned. A lot of them are creating task forces to look into issues like how do you spot AI manipulation and scale back some of the misinformation disinformation efforts propagated by AI. We're going to keep looking at how this plays out in congressional chambers, in courtrooms. We'll be examining efforts to combat AI enabled misinformation. So, we'll look at the content provenance debate. We'll look at the labeling issue, and we're hoping to talk a little bit more about media literacy. What can people do to educate themselves to be prepared for the coming rush of AI generated content?

Mr. Schwartz: Do you see any legislative action coming about media literacy?

Ms. Hsu: There are certain states that are attempting to push through. Really, they're media literacy guidelines, they're not laws for the most part. It's one of those things that legislators realize should be a priority, but it still really isn't.

Mr. Schwartz: Tiffany, thank you so much for helping us understand this really complex issue. Really appreciate it.

Ms. Hsu: Thanks for having me on to discuss this very depressing topic.

Outro: If you enjoyed this podcast, check out our larger suite of CSIS podcasts from Into Africa, the Asia Chessboard, China Power AIDS 2020. The trade guys, Smart Women, Smart Power, and more. You can listen to them all on major streaming platforms like iTunes and Spotify. Visit csis.org/podcasts to see our full catalog.