#1614 Deep-Fakery and Deep Consequences for Democracy

JAY TOMLINSON - HOST, BEST OF THE LEFT: [00:00:00] Welcome to this episode of the award winning *Best of the Left* podcast in which we grapple with the fact that AI generated deepfakes, entirely fabricated audio and video of recognizable people, are here. They have been on the horizon for years, but they have finally arrived during the biggest global election year in history, which may prove to be a make or break year for democracy itself, as we struggle to separate fact from fiction and autocracy is on the rise around the world. Sources today include *Forbes*, *CYBER*, *All Things Considered*, *Aperture, On with Kara Swisher,* and *TED Talks Daily,* with additional members-only clips from *Forbes* and *What Next: TBD.*

Deepfaking Democracy: Why AI Threatens News And Global Elections In 2024 Part 1 - Forbes - Air Date 2-6-24

UNIDENTIFIED NARRATOR: 2024 will be a record year for elections around the world. Over 4 billion people, more than half of Earth's population, are expected to cast a ballot. 7 out of 10 of the most populous nations are going to the polls. [00:01:00] And many elections will be in countries consequential to the news cycle.

Taiwan held its presidential elections on January 13th, which saw William Lai of the Democratic Progressive Party win with over 40 percent of the vote. Lai's election is expected to make relations between Taiwan and mainland China more antagonistic. Both Ukraine and Russia, who remain locked in war with one another, have scheduled elections in March and U. S. elections in November are bound to draw intense international attention in what is shaping up to be a rematch of the 2020 elections.

PRESIDENT JOE BIDEN: Democracy is still at risk. This is not hyperbole.

UNIDENTIFIED NARRATOR: Many academics, political analysts, and think tanks expect 2024 to be a major stress test on the concept of democracy itself. And one particular variable that will further complicate this test is the rise of AI tools, and the ability to create convincing, deepfake news content.

JORDAN PEELE AS VOICE OF PRESIDENT OBAMA DEEPFAKE: We're entering an era in which our enemies can make it look like anyone is saying anything at any [00:02:00] point in time, even if they would never say those things. Moving forward, we need to be more vigilant with what we trust from the internet. It's a time when we need to rely on trusted news sources.

ALEXANDRA S. LEVINE, FORBES WRITER: deepfakes and cheap fakes are not new. But with the explosion of AI that was ushered in by the introduction of ChatGPT just over a year ago, we saw deepfakes proliferate. And the types of deepfakes that we've been looking at, which are these fake news segments using the real likeness and real logos of real news outlets and the faces of real broadcasters, are seemingly new and they are particularly problematic right now as we are heading into a really high stakes election and also as we are in the midst of a war.

We have seen deepfake news segments from top prominent anchors at all sorts of outlets ranging from CNN to CBS and beyond.

YENA LEE, FRANCE 24 NEWS ANCHOR: Truth or fake? You're beginning with a story of a [00:03:00] video on social media where President Zelensky appears to surrender to Russian forces. What's that about?

CATALINA MARCHANT DE ABREU, FRANCE 24 CORRESPONDENT: A false video of President Zelensky was diffused yesterday where he's apparently making an announcement giving up to Russian forces. This video was diffused on a hacked Ukrainian news website called Ukraine 24.

HANY FARID: I've been seeing it sort of come in and out for several years now, but in really seeing it consistently in high quality, I would say in the last 12 months.

DEEPFAKED PRESIDENT ZELENSKY: I have to make difficult decisions. At first, I decided to return Donbas. It's time to look in the eye.

HANY FARID: There's two main reasons for it. One is that the technology to create deepfakes of news anchors has just gotten better. But two, and I think this is also important, is that most of the major social media companies have eviscerated their trust and safety teams. And that's not just Twitter, by the way. That one's easy. But it's even the Facebooks of the world, the YouTubes, and the TikToks. [00:04:00] And so, as a result of that, when people create fake content, it's much, much easier to distribute.

So, remember that when we're talking about deepfakes, there's really three parts to it. There's the underlying technology, the bad actors who are misusing these technologies, but then there's the spread of that. And the spread of that technology is not an AI question, that's a social media question. All three things have now lined up. The technology is getting better, bad actors are figuring out that you can monetize or abuse this content, and the social media companies have fallen asleep at the wheel again.

BILL WHITAKER, CBS ANCHOR: Using video from the *CBS News* archives, Chris Ume was able to train his computer to learn every aspect of my face and wipe away the decades. This is how I looked 30 years ago. He can even remove my mustache.

HANY FARID: There are two approaches to detecting manipulated media, what we call proactive and reactive. So, the reactive is sort of my bread and butter here as an academic at UC [00:05:00] Berkeley. What we do is we take an image, an audio or a video, we run it through a battery of tests, and we try to figure out if it's been manipulated or AI generated, all after the fact, right? So, stuff gets online, some fact checker contacts us, we analyze the content, and we eventually tell the fact checker and they eventually set the record straight, and meanwhile the whole world has moved on and gotten defrauded to the tune of millions of dollars.

So, the reactive stuff is good, if you will, as a post-mortem, but at the speed at which social media moves, half life of a social media post can be measured in minutes, you're not there fast enough to deal with the damage. The proactive techniques, the way they work, is that if you pick up your phone and record something, or you are in the business of generating AI content, you can inject into that content, whether it's real or AI generated, a digital watermark that is cryptographically signed and then downstream your browser or a piece of software can read that watermark and say, Nope, I know that this is AI generated, or in [00:06:00] fact that it is real, and you can do that instantaneously.

This only works when you have good players. So, when Adobe decides it's going to put watermarks into its content, well great, I trust Adobe. But, a lot of bad players out there, and a lot of this code for creating deepfakes is open source. So if you have open source, and you've got some code in there for inserting a watermark, well the bad guy's going to go in there and remove that code, and we're off to the races.

So, the watermarking absolutely are going to play a role here, but they will not, in and of themselves, solve the problem because there's always ways around this

technology and there's open source and there's bad actors. But I'm super supportive of that for the big players like Adobe, OpenAI, and MidJourney, and maybe we lop off half the problem.

AI Deepfakes Are Everywhere and Congress is Completely Out of Their Depth - CYBER - Air Date 2-9-24

LIA HOLLAND: This is incredibly complicated, but one of the places to start is with the existing laws that a bunch of these people who've been affected that we're already talking about, are already suing under. Most states have a right of publicity law that allows, you know, celebrities or public figures to [00:07:00] sue if people misuse their images in some sort of manner that is commercial or could be construed as commercial. And then at the same time, we have defamation law, which can often be a way for average people to sue those who humiliate them, while it protects celebrities less.

So, between those two, depending on where you live, and in the majority of states, there's some good stuff there. There's some good mechanisms, at least in terms of if you want to get a lawyer and if you want to sue the person who's causing you misery. But still that doesn't actually give victims of these deepfakes a real time way to say, Hey, get this disgusting porn of me off the internet. This is humiliating me and it's spreading everywhere. And while it pains me, because I know how extensively something like the Digital Millennium Copyright Act has been abused, these sort of notice and take down systems, I'm talking to [00:08:00] a lot of people and none of us really see another way to make something that is actually responsive to the harms that people are going to be experiencing. But what we can get right this time, and I think, you know, with Donald Trump right now claiming that actual photos of him are deepfakes is that unlike the, we could build a system where there are actual consequences if you abuse it, if you say that's an embarrassing video of you or a video of police misconduct or what have you is a deepfake and it actually isn't. So that's where I'd start.

MATTHEW GAULT - HOST, CYBER: Oh, that's interesting. I didn't even think about the possibility of a public person getting into trouble for lying about a real image being a deepfake.

LIA HOLLAND: Oh, yeah, that's coming.

MATTHEW GAULT - HOST, CYBER: Has anyone proposed legislation? Or is this just like conversations?

LIA HOLLAND: This is conversations because I think that a lot of the people who are looking at No Fakes or No AI Fraud or what have you and saying that these are [00:09:00] terrible laws with unintended, extremely harmful consequences are also really feeling for the reality that people are going to face with these technologies and knowing that we need to do something. And happily, we're having much more of a proactive conversation here amongst ourselves about what we do want, then I think we have maybe with previous online revolutions.

JANUS ROSE: That was one of the things that struck me when initially I wrote about the No Fakes and then there was another law in Tennessee that I think was proposed the same day, which is that, you know, and this is like what I wrote about in the article I wrote a couple of weeks ago, which is like, it seems like the gist of these laws is sort of intended to protect celebrities, and maybe the rest of us, too, kind of, sort of? And that's kind of like where I came at this from, which is like the fact that like, you know, a lot of people have been talking about this as a concern for a while, but now that Taylor Swift is mad, now that, like, The Weeknd and Drake are mad about this... and it's not just, you [00:10:00] know, like photos. It's also, like, voice rights, music, and then all other kinds of stuff that could be considered intellectual property or personalized, sort of like, intimate representations of someone's person.

It seems to me like that's where this always starts and ends, when we get "privacy protections" or something that is supposed to at least in theory be like protecting privacy, is that it's generally winds up protecting famous and rich people and doesn't do a whole lot for regular people who are facing abuse and harassment and, you know, sexual violence. And the copyright system, as you were just mentioning, like the copyright system, you're saying, like, Oh, we keep looking at this, like, we don't see any other way of enforcing that. So, like, what what would be different about this compared to, like, you know, the fact that when artists, for example, have people profiting from their art that they release, [00:11:00] like music or otherwise, there is a mode for redress, but it's not very accessible unless you have a lot of money to litigate it. So, what's the fix here when it comes to this stuff, if in the past, this has been kind of the status quo?

LIA HOLLAND: That's a great question, uh, because we are really swimming upstream against the headline grabbers here. If something horrible is being done to Taylor Swift and, you know, by God, Josh Hawley can trot out there with a bill and wave it in the air and it's going to get covered, you know, all over

creation, the motive there is really clear and straightforward. Politicians like laws that grab headlines. They like flashy partnerships with celebrities. And I would also say that the lobby of those IP rights holders, the major labels and publishers and content companies and what have you, is extremely powerful and really well organized. And from the moment that this all blew up, they've been in legislative offices, you know, gunning for a bill that's going to benefit, you know, the Universal Music Groups of the world.

And yeah, and that's [00:12:00] why, for myself, I turn towards - and I think that there are also legislators who are thinking in this way - better tools for everyday people, because I don't think that there's going to be an effective way to censor, or we can't put the rabbit back in the hat with AI. What we need is proactive tools to address it in a way that is, you know, minimally invasive when it comes to surveillance and censoring speech, and slapping upload filters across the whole Internet isn't the right thing either. I would look at something like, Well, we've got Google reverse image search and we know that that works pretty good. And we've got, you know, the, the DMCA and that, you know, there's an established protocol for that. And we've got this idea that... and we've learned a lot since since that legislation went in. And so, can we slap something together that doesn't reinvent the wheel and just gives people the right to say, Hey, this is [00:13:00] a horrific fake photo of me, please scrub it off the internet. And they can make that request in a way that the platforms have to be accountable to. Cause I think that's the other thing. It's really hard to be heard as an individual when platforms are dealing with so many users.

JANUS ROSE: About that point on platforms being accountable: this is kind of like something I say a lot when I'm talking about this topic, is that like, we're kind of addressing a symptom of a problem here and not the actual problem. And the problem is that we have all these giant tech companies that are producing this technology and they basically have no regulation and they're kind of just doing whatever they want. And that's, you know, there's even, these lobbying groups, like, Elon Musk has this AI institute that's essentially saying, You must let us develop AI and if you don't, then you're killing people. People will die. That's kind of, like, what I always frame this around is, we're dealing with a symptom and not the actual problem, which is that, even the way in which these tech companies address [00:14:00] this problem sometimes is very much like Band-Aid oriented. I was writing an article a couple of weeks ago about the filtering system on some of these things. OpenAI has been constantly needing to patch ChatGPT and Dall-E and all these image generators, because people keep finding ways to get past the content filter system that prevents them from generating certain types of images through all these kind of tricky ways. And it's just this cat and mouse game. And, you know, when it comes to some of these - I was reading this paper - and when it comes to some of these systems,

what they're actually doing is that they're still generating the content and then they're just not showing it. So, it's not even that they're preventing the content from being created in the first place. They're just filtering it out.

MATTHEW GAULT - HOST, CYBER: But really, it's off-camera sketching that image you asked for and just storing it in a digital warehouse somewhere and all the forbidden images you'll never be able to say.

JANUS ROSE: Yeah, but it's like, even if nobody sees that image, [00:15:00] that's indicative of a larger problem, which is that we don't actually know how to stop that from happening, because the training has already occurred. These systems are already built on top of billions of images that were taken without permission. And, you know, some of them - we wrote another story about this a couple of weeks ago - LAION, which is a probably one of the most commonly used image databases used in generative AI, it contains billions of images that are taken from web scraping and it was found that I think about 3000 instances of CSAM, of child exploitation, were found in this massive database. That's kind of an example of what we're dealing with here. It's like, the sort of, like, base problem has already occurred. We're just getting the results of it now and you can filter the results, but that doesn't ultimately solve the fact that where this all came from.

Tech giants pledge action against deceptive AI in elections - All Things Considered - Air Date 2-16-24

JUANA SUMMERS - HOST, ALL THINGS CONSIDERED: More than 40 countries are set to hold major elections this year, and many experts worry that rapidly evolving artificial intelligence technologies could [00:16:00] disrupt those votes. Just a few weeks ago, an apparent deepfake robocall that sounded like President Joe Biden told people not to vote in New Hampshire. Today, 20 major tech companies announced they are going to do their part to avoid becoming the story.

Joining us now to talk through this new agreement are NPR's Shannon Bond, who covers how information travels, and Miles Parks, who covers voting. Tell us about this agreement. What's in it?

SHANNON BOND: Well, it's aimed at AI-generated images, audio and video that could deceive voters, so whether that's by impersonating a candidate doing or saying something they didn't or misleading people about, you know, when or

how to vote. And the companies are agreeing to some pretty broad commitments here to develop technology to watermark AI content and to detect and label these kind of fakes. They're pledging to be more transparent about how their tools and platforms are being used. They want to educate the public about AI.

Now, look. Many of these actions are things some of these companies are already working on. And what's notable [00:17:00] here is that this agreement does not outright ban this kind of deceptive use of AI in elections.

JUANA SUMMERS - HOST, ALL THINGS CONSIDERED: Right. OK. Let's dig in a little bit here. Does this agreement actually bind these companies to do anything or is this more of like a mission statement?

SHANNON BOND: Yeah. This is a voluntary agreement, so it's not binding. And remember, just because companies create policies about AI doesn't mean they always effectively enforce them. Now, this agreement came together in just the past six weeks. And in many ways, you know, it seems like it had to be pretty broad to get this many companies to agree. We spoke with Microsoft President Brad Smith today. He said that unity itself is an accomplishment.

BRAD SMITH: We all want and need to innovate. We want and need to compete with each other. But it's also just indispensable that we acknowledge and address the problems that are very real, including to democracy.

SHANNON BOND: And indeed, even as these companies, including Microsoft, are saying, you know, they're on guard over risks of AI, they're also continuing to roll out even more [00:18:00] advanced technology. Like, just yesterday, OpenAI, one of the other companies that signed this agreement, they announced this tool that allows you to type in a simple text description to create a really realistic high-definition video.

JUANA SUMMERS - HOST, ALL THINGS CONSIDERED: I mean, hearing you describe that, it's easy to see how a tool like that could be used to spread lies about voting, for example. Miles, over to you. How are elections officials feeling about AI right now?

MILES PARKS: They are thinking about it a lot. Last week, I was at a conference with some of the top election officials in the country. They don't want people to panic. Generally, they see AI as more of an extension of problems they were already working on. That's how Adrian Fontes, who's the secretary of state of Arizona, that's how he put it to me when we were talking.

ADRIAN FONTES: AI needs to be demystified. AI needs to be exposed for the amplifier that it is, not the great, mysterious, world changing, calamity inducing, you know, monstrosity that some people are making it out to be.

MILES PARKS: That said, there are a myriad of ways experts can imagine these tools threatening democracy even beyond, I think, the most obvious use [00:19:00] case, which is, you know, making a fake video of a candidate saying something they didn't actually say.

JUANA SUMMERS - HOST, ALL THINGS CONSIDERED: Walk us through, if you can, some of those scenarios.

MILES PARKS: Yeah. I asked Smith from Microsoft about this, And he said specifically he's worried about people using AI to dub over real videos with fake audio. That could be a lot more convincing to people than creating a whole new video. But there's also a bigger picture worry that I heard percolating at this conference last week, that as more fake stuff is swirling online, the public will slowly lose trust in all information. That's one of the hardest aspects of this accord.

The tech companies say they want the public to be more skeptical of what they see online, but that can lead to this feeling among people that nothing is true or real. And bad actors can capitalize on that too, by then being able to claim that real information is fake. It's called the liar's dividend. With more AI-generated stuff floating around, it's just going to become more and more common that candidates when real bad information comes out about them, they can just say, no, that's fake. That's AI-generated.

JUANA SUMMERS - HOST, ALL THINGS CONSIDERED: I mean, we should just point out here that policing [00:20:00] truth and lies online is really fraught these days. The political right in particular has cast these kinds of efforts as politically biased. Are tech companies worried about diving in here?

SHANNON BOND: Yeah. I asked Brad Smith of Microsoft about this. You know, he and other tech executives involved in this, they say there is a clear distinction here between free expression, which they say they're all committed to, and using AI or other kinds of technology, you know, in a way that is really deceiving, misleading voters, interfering with the election process. They're very much framing this fight as one against fraud.

MILES PARKS: I do think that we'll probably see companies jump in a lot harder against things that are explicit lies about how people vote. Think of like a

video that claims Election Day is on Friday versus Tuesday. That's pretty easy to police. I think it's the content that raises doubts about the trustworthiness about elections, that it's still an open question how companies are going to police that sort of content in 2024.

Deepfake Adult Content Is a Serious and Terrifying Issue - Aperture - Air Date 5-1-23

MIKE MCEWEN - HOST, APERTURE: As of 2019, 96 percent of deepfakes on the [00:21:00] internet were sexual in nature, and virtually all of those were of non-consenting women. With the release of AI tools like DALL·E and Midjourney, making these deepfakes has become easier than ever before, and the repercussions for the women involved are much more devastating.

Recently, a teacher in a small town in the United States was fired after her likeness appeared in an adult video. Parents of the students found the video and made it clear they didn't want this woman teaching their kids. She was immediately dismissed from her position.

But this woman never actually filmed an explicit video. Generative AI created a likeness of her and deepfaked it onto the body of an adult film actress. She pleaded her innocence, but the parents of the students couldn't wrap their heads around how a video like this could be faked. They refused to believe her. And honestly, it's hard to blame them. We've all seen just how good Generative AI can be. This incident, and many others just like it, proved how dangerous AI adult content is and, if left unchecked, it could be so, so much worse.

[00:22:00] At first glance, AI pornography might seem harmless, if we can generate other forms of content without human actors, why not this one? Surely it may reduce work in the field, but it could also curb more problematic issues in the industry. If the AI was used to create artificial people, it wouldn't be so bad, but the problem is that the generative AI has been mainly used with deepfakes to convince viewers that the person they're watching is a specific, real person, someone who never consented to be in the video.

Speaking of consent, by convincingly portraying women in suggestive situations, the perpetrators commit sexual acts or behaviors without the victim's permission, and that, by definition, is sexual assault. But does using generative AI to produce these videos cause any actual harm beyond being defined as

assault? For the victims involved, there are numerous consequences to being portrayed in these videos.

QTCINDERELLA: This is what it looks like to see yourself naked against your will being spread all over the internet.

MIKE MCEWEN - HOST, APERTURE: QTCinderella is a Twitch streamer who built a massive following for her gaming, baking, and lifestyle content. [00:23:00] She also created the Streamer Awards to honor her fellow content creators, one of whom was Brandon Ewing, aka Atrioc. In January of 2023, Atrioc was live streaming when his viewers saw a tab open on his browser for a deepfake website. After getting screenshotted and posted on Reddit, users found that the site address featured deepfakes videos of streamers like QTCinderella doing explicit sexual acts.

Cinderella began getting harassed by these images and videos and after seeing them she said, "The amount of body dysmorphia I've experienced seeing those photos has ruined me. It's not as simple as just being violated, it's so much more than that". For months afterwards, QTC Cinderella was constantly harassed with these reminders of these images and videos. Some horrible people sent the photos to her 17 year old cousin.

And this isn't a one off case. Perpetrators of deepfakes are known to send these videos to family members of the victims, especially if they don't like what the victim is doing publicly. The founder of Not Your Porn, a group dedicated to removing non-consensual porn from [00:24:00] the internet, was targeted by internet trolls using AI generated videos, depicting her in explicit acts. Then, somebody sent these videos to her family members. Just imagine how terrible that must feel for her and her relatives.

The sad truth is that even when a victim can discredit the videos, the harm might already be done. A deepfake can hurt someone's career at a pivotal moment. Cinderella was able to get back on her feet and retain her following, but the school teacher, who lost her livelihood, wasn't so lucky. Imagine someone running for office and leading in the polls, only to be targeted with a deepfake video 24 hours before election night. Imagine how much damage could be done before their team could prove that the video was doctored.

Unfortunately, there's very little legislation on deepfakes, and so far, only three states in the US have passed laws to address them directly. Even with these laws, the technology makes it difficult to track down the people who create them. Also, because most of them post on their personal websites rather than on

social media, there's no regulations or content moderation limits on what they can share. Since [00:25:00] tracking and prosecuting the individuals who make this kind of content is so challenging, the onus should be on the companies that make these tools to prevent them from being used for evil.

And in fairness, some of them are trying. Platforms like DALL·E and Midjourney have taken steps to prevent people from creating the likeness of a living person. Reddit is also working to improve its AI detection system and has already made considerable strides in prohibiting this content on its platform. These efforts are important, but I'm not sure they'll completely eliminate the threat of deepfakes. More generative AI tools are coming on the scene and will require new moderation efforts, and eventually some of these platforms won't care, especially if that gives them an edge over well established platforms.

And then there's the sheer influx of uploaded content. In 2022, Pornhub received over 2 million video uploads to its site. That number will likely increase with new AI tools that can generate content without needing a physical camera. How can any moderation system keep up with that insane volume?

The worst thing about these deepfakes is that the victims can't just log off of the internet either. Almost all of our [00:26:00] livelihoods depend on the internet, so logging off would be an enormous disadvantage in their careers and personal life. And expecting anyone to leave the internet to protect themselves isn't a reasonable ask. The onus isn't on the victim to change, it's on the platforms and the government to create tools that prevent these things from happening so easily. If all the women who are being harassed went offline, the trolls would win, and this tactic of theirs would be incredibly successful. They could effectively silence critics and whoever they felt like attacking.

There is another problem with generative AI tools producing so much adult content. It introduces strong biases to the algorithms and how women should be presented. Many women have reported that they're often over sexualized when they try to create an image of themselves using AI tools. These biases are introduced by the source of the AI's training data, the internet. Although nudes and explicit images have been filtered out for some generative AI platforms, these biases still persist. These platforms have to do more than just let the open internet train their AI if they want to prevent the overt sexualization of women to be their normal output.

[00:27:00] Deepfakes may be making headlines now, but the truth is they've been around in spirit for a very long time. Before generative AI, people used tools like Photoshop and video editing software to superimpose celebrities heads on the bodies of adult film actors. Broadly, these doctored videos weren't

compelling, but the things are now very different with AI. We're careening dangerously close to a point where we can no longer discern the real from the fake. French postmodern philosopher Baudrillard warned of a moment when we can no longer distinguish between reality and a simulation. Humans use technology to navigate a complex reality. We invented maps to guide us through an intricate mass of land. Eventually, we created mass media to understand the world around us and help simplify its complexity. But there will be a point where we lose track of reality, a point where we're spending more time looking at a simulation of the world on our phone than we will be participating in the real world around us, and we're almost there now.

With generative AI, our connection to reality is even further disconnected, because technology can convincingly replicate reality on our [00:28:00] devices, we're less inclined to go outside and see what's real for ourselves. This inability of human consciousness to distinguish what is real and what is simulation is what Baudrillard called hyperreality. A state that leaves us vulnerable to malicious manipulation, from things like deepfakes to people getting fired, to propaganda leading to the loss of millions of lives. You might remember that a couple of years ago there were numerous PSAs, often from celebrities warning us to keep an eye out for deepfakes. They were annoying, but ultimately, they succeeded in making the public hyper aware of fake videos. But not so much with the deepfake adult content. Maybe it's because the PSAs about deepfakes didn't mention pornography, they addressed fake speeches by presidents and famous people instead. Or maybe it's because those who consume this content don't care whether it's real or fake. They're okay with the illusion. One thing is true though, if the general public was trained to recognize deepfake pornography, the potential for harm would be limited. By being more critical as information consumers and reporting these harmful videos when we see them, we might be [00:29:00] able to curb the effects of this dangerous new medium.

It's not like we're strangers to being critical of what we see and read online. When Wikipedia was first introduced, the idea that it could be a legitimate source of information was laughable. It was mocked on sitcoms and late night television, it symbolized the absurdity of believing what you read on the internet. That perception changed with time, deservedly so for Wikipedia, but we had a healthy skepticism towards user generated internet platforms for a while. The question is can we be critical and discerning towards deepfakes while acknowledging that some content is real? Will we lose track of what's simulation and what's reality and just distrust whatever we see online? Or worse, will manipulators succeed in making deepfake inflicted suffering an everyday occurrence, and we end up accepting that as the cost of existing

online? And is there any hope of regulation stopping the constant assault of generative AI on our well being?

Will Killing Section 230 Kill the Internet? - On with Kara Swisher - Air Date 2-23-23

EVELYN DOUEK: I think that there are real legitimate questions about the breadth of 230 is the way the lower courts have interpreted it. I think, you know, Hany talked about the Snapchat case earlier, which is [00:30:00] a good example of where 230 immunity was pierced. And I think, you know, there are other really good questions around really bad actor platforms that know all of this stuff is going on and not taking action.

KARA SWISHER - HOST, ON WITH KARA SWISHER: Team mental health, for example?

EVELYN DOUEK: Yeah. I mean, I think, you know, there's going to be causal chain problems on some of those cases, but, you know, I do think that Hany's absolutely right. The court took these cases because there's sort of hunger, that's, Everyone's talking about section 230. We should be talking about section 230. But I think that these weren't the fact sets that they thought. And so it'll be interesting to see if they come back and have another bite at it soon.

KARA SWISHER - HOST, ON WITH KARA SWISHER: Jeffrey, is there another case?

JEFFREY ROSEN: Well, the ones we've talked about from Florida and Texas, which, as everyone said, the court will take next year, involve a different question about the scope of 230, but one that the court is likely to divide over, and it's possible that that could have implications for how liability is applied in other cases too. But that's going to be absolutely fascinating and so squarely poses the conflict about whether or not the platform should be treated as common carriers and obey First Amendment standards and in some ways, those will even [00:31:00] be more constitutionally significant than these cases.

KARA SWISHER - HOST, ON WITH KARA SWISHER: All right. Is there any other industry that gets blanket immunity protections the way social media companies do? Everybody gets sued, except them. Is there any sort of parallel here? Can any of you think?

HANY FARID: No, there isn't. I mean, I'm not the legal scholar here, but we've heard this, and I think even one of the justices says is during the Gonzalez hearing is why does the tech industry get so much protection? Every other industry has to internalize these risks and deal with it. And I don't know of any other industry that has this type of almost blanket immunity.

EVELYN DOUEK: I mean, you know, the tech industry obviously gets sued all the time, but I do think that there, I mean, this is a somewhat exceptional statute provided for what Congress recognized at the time as an exceptional situation, which is, you know, these platforms have been the become the custodians of all of our speech. And I think, you know, the important thing to remember at section 230 is, yes, it provides platforms immunity but it also provides users immunity and the point of that platform immunity is to [00:32:00] protect the speech of users. I'm sounding much more libertarian on this podcast than I intended to, I have to say. You know, I really do think...

KARA SWISHER - HOST, ON WITH KARA SWISHER: That's alright. You've lived in Silicon Valley.

EVELYN DOUEK: Yes, six months. That's all it took. There's something in the water.

KARA SWISHER - HOST, ON WITH KARA SWISHER: You can be libertarian-light, which is most of them, honestly. They call themselves that.

EVELYN DOUEK: I think content moderation is extremely important. I just get nervous about government rules that incentivize overmoderation and that platforms that don't care about sort of marginalized communities or disparate impacts end up, you know, we have seen this before with sort of the Foster amendments as well, taking down speech of people who, you know, don't have the same resources. So.

HANY FARID: Can I follow up on that, Kara? So. Evelyn raises an absolutely valid point that we do have to be careful about overmoderation. I will point out, however, that when we passed the DMCA, the Digital Millennial [sic] Copyright Act, these same claims were being made by the tech companies that you are going to force us to over moderate to avoid a liability, and it wasn't true. And look, DMCA is not perfect, but it has largely been fairly effective and [00:33:00] it created a healthy online ecosystem that has allowed us now, for both creators and producers, to monetize, music and movies and art. And so when you have rules of the road, they can actually be very, very good at creating a healthier online ecosystem. And since the companies are incentivized

to keep content up, that's the financial side, I think that on balance, this might actually work out even if there is more liability with reduction of 230 protection.

JEFFREY ROSEN: I would just say that industries that are immunized from suits include lawyers, the ones who are most protected and all the privileges that the courts have protected against ineffective assistance of counsel claims or the lawyer-client privilege, are designed to protect deliberative privilege and First Amendment values; the same with executive privilege, when you can't sue the executive to get the deliberations so that you can get. advice. So, this immunity, as Evelyn [00:34:00] says, for the platforms is designed to achieve a First Amendment value, which is, deliberation and not overmoderating. And it's heartening, despite the really tough questions that are on the horizon involving the scope of the First Amendment, to see a consensus that 230 did achieve its purpose. And there's a reason that the US has a freer free speech platform than Europe, for example, which lacks this immunity, and the consequences of abandoning it might be severe. So, let's just pause during this brief moment of agreement, not to sing Kumbaya, but to say it's great that thinking about this hard, the justices may be inclined to think that 230 isn't so bad after all.

KARA SWISHER - HOST, ON WITH KARA SWISHER: So, my last question because, that you led me perfectly to it. There's two ways to go here, is that, you know the swirl and how powerful these social media companies are. There's one way where Google, Twitter, Meta, et cetera, gets their ships in order without legislative or judicial action because they should be in charge of all this stuff because they were duly elected by nobody. Or, as Kagan [00:35:00] specifically called out Congress to act, which are our elected officials, as damaged as they may be. Two things: one, who should be running the show here? And let's imagine a world with rational internet regulations, what would those be and what would the internet look like? Hany, you start with the first one and then Jeffrey and Evelyn you can answer the second one

HANY FARID: There is no evidence that the technology company can self-regulate. The last 25 years has taught us this. And not only that is that the business model that has led to the Googles and the Facebooks and the TikToks of the world continues to be the dominant business model of the Internet, which is engagement driven, ad driven, outrage driving. And that business model is the underlying root poison, I would argue. I don't think we can sit around and wait for the companies to do better. I don't think they will. There is no evidence of it. I think despite the fact that I don't want the regulators putting rules of the road, I think there is no other choice here. Ideally, by the way, the [00:36:00] consumers would have made the choice. We would have said, okay, we don't like the way you're doing business, we're going to go elsewhere. But in addition

to phenomenal wealth they have virtual monopolies and so we as the consumer don't even have choices and that means the capitalism won't work here. And so we need the regular regulators to step in.

KARA SWISHER - HOST, ON WITH KARA SWISHER: All right, Jeffrey. Congress should act? My feeling is Congress should have done privacy and antitrust legislation and taken care of this in a whole different way. But, what do you think about that part?

JEFFREY ROSEN: I guess the quick question first is, will it act and what should it do? And will it? Probably not, because there's not consensus as we've been discussing with conservatives more concerned about content discrimination, for better or for worse, and liberals more concerned about hate speech and harmful conduct. I find it hard to imagine what a national free speech regulation would look like. And in fact, I can't imagine one that's consistent with First Amendment values short of imposing them, which there's an argument for not doing at the federal level because companies need some play in the [00:37:00] joints to take down some more offensive speech than the First Amendment protects, while broadly allowing a thousand flowers to bloom.

The one interesting consequence of this argument is to make me think, you know, the companies, although it's messy and there's lots to object to, it may be better than the alternatives of either really sweeping, imposing a First Amendment standard on the federal level or allowing a great deal more moderation than would be consistent with First Amendment values.

KARA SWISHER - HOST, ON WITH KARA SWISHER: Evelyn, you get the last word. 230 looks like it's going to live to fight another day.

EVELYN DOUEK: Yeah. There is no rational world where the best way to make tech policy is by nine, you know, uh, older justices weighing in on a case every 20 something years to sort of catch up on what's been going on. That is not how this should happen. Absolutely, Congress, you know, if it could get It's act together it could pass some legislation enabling a digital agency that could be even more nimble, and sort of, you know, gather facts and understanding in which to make [00:38:00] policy that's more sort of finally attuned to the problem. And, you know, then we could talk. Absolutely, Kara.

You know, you mentioned privacy and antitrust, that would be a hundred percent the sort of place where I would start. I would also really start on transparency legislation and data access. You know, what are these platforms doing and are they doing what they say they're doing? Let's get researchers in.

And that's where I'd start. Cause you can't solve problems that you don't understand. And I think that that's step one. And the only other thing, you know, before we close, this has been a very sort of parochial conversation, but there are other legislatures, and Europe is taking action. The Digital Services Act is coming, and so these platforms are going to have to change and adjust anyway, because they're going to be regulated, you know, no matter what the Supreme Court does.

When AI can fake reality, who can you Trust? | Sam Gregory - TED Talks Daily - Air Date 12-20-23

SAM GREGORY: The last thing we need is a diminishing baseline of the shared, trustworthy information upon which democracies thrive, where the specter of AI is used to plausibly believe things you want to believe and plausibly deny things you want to ignore. But I think there's a way we can prevent that future, if we act now; that if we prepare, don't panic, [00:39:00] we'll kind of make our way through this, somehow. Panic won't serve us well, [it] plays into the hands of governments and corporations who will abuse our fears, and into the hands of people who want a fog of confusion and will use AI as an excuse.

How many of you know someone who's been scammed by an audio that sounds like their kid? And for those of you who are thinking, I wasn't taken in. I know how to spot a deepfake, any tip you know now is already outdated. Deepfakes didn't blink. They do now. Six fingered hands were more common in deepfake land than real life. Not so much. Technical advances erase those visible and audible clues that we so desperately want to hang on to as proof we can discern real from fake.

But it also really shouldn't be on us to make that guess without any help. Between real deepfakes and claimed deepfakes, we need big picture structural solutions. We need robust [00:40:00] foundations that enable us to discern authentic from simulated, tools to fortify the credibility of critical voices and images, and powerful detection technology that doesn't raise more doubts than it fixes. There are three steps we need to take to get to that future.

Step one is to ensure that the detection skills and tools are in the hands of the people who need them. I've talked to hundreds of journalists, community leaders, and human rights defenders, and they're in the same boat as you and me and us. They're listening really closely to the audio, trying to think, can I spot a

glitch? Looking at the image, saying, Ooh, does that look right or not? Or maybe they're going online to find a detector, and the detector they find they don't know whether they're getting a false positive, a false negative, or a reliable result.

Here's an example. I used a detector which got the 'pope in the puffer jacket' right, but then when I put in the Easter Bunny image that I made for my kids, it said that it was human generated. This is because of some big [00:41:00] challenges in deepfake detection. Detection tools often only work on one single way to make a deepfake, so you need multiple tools. And they don't work well on low quality social media content. Confidence score; how do you know whether that's reliable? If you don't know if the underlying technology is reliable, or whether it works on the manipulation that has been used. And tools to spot an AI manipulation don't spot a manual edit.

These tools also won't be available to everyone. There's a trade off between security and access, which means if we make them available to anyone, they become useless to everybody. Because the people designing the new deception techniques will test them on the publicly available detectors and evade them.

But we do need to make sure these are available to the journalists, the community leaders, the election officials globally, who are our first line of defense, thought through with attention to real world accessibility and use. Though, at the best [00:42:00] circumstances, detection tools will be 85 to 90 percent effective, they have to be in the hands of that first line of defense. And they're not right now.

So for step one, I've been talking about detection after the fact. Step two: AI is going to be everywhere in our communication. Creating, changing, editing. It's not going to be a simple binary of, Yes, it's AI, or, Phew, it's not. AI is part of all of our communication. So we need to better understand the recipe of what we're consuming. Some people call this content provenance and disclosure. Technologists have been building ways to add invisible watermarking to AI generated media. They've also been designing ways, and I've been part of these efforts within a standard called the C2PA, to add cryptographically signed metadata to files. This means data that provides details about the content cryptographically signed in a way that reinforces our trust in that information. It's an [00:43:00] updating record of how AI was used to create or edit it, where humans and other technologies were involved, and how it was distributed. It's basically a recipe and serving instructions for the mix of AI and human that's in what you're seeing and hearing. And it's a critical part of a new AI-infused media literacy.

And this actually shouldn't sound that crazy. Our communication is moving in this direction already. If you're like me, you can admit it, you browse your TikTok 'For You' page, and you're used to seeing videos that have an audio source, an AI filter, a green screen, a background, a stitch with another edit. This, in some sense, is the alpha version of this transparency in some of the major platforms we use today. It's just that it does not yet travel across the internet, it's not reliable, it's not updatable, and it's not secure.

Now, there are also big challenges in this type of infrastructure for authenticity. As we create these durable [00:44:00] signs of how AI and human were mixed, that carry across the trajectory of how media is made, we need to ensure they don't compromise privacy, or backfire globally. We have to get this right. We can't oblige a citizen journalist filming in a repressive context, or a satirical maker using novel gen AI tools to parody the powerful, to have to disclose their identity or personally identifiable information in order to use their camera or ChatGPT. Because it's important they be able to retain their ability to have anonymity at the same time as the tool to create is transparent. This needs to be about the how of AI human media making, not the who.

This brings me to the final step. None of this works without a pipeline of responsibility that runs from the foundation models and the open source projects through to the way that is deployed into systems, APIs, and apps to the platforms [00:45:00] where we consume media and communicate.

I've spent much of the last 15 years fighting essentially a rearguard action like so many of my colleagues in the human rights world against the failures of social media. We can't make those mistakes again in this next generation of technology. What this means is that governments need to ensure that within this pipeline of responsibility for AI, there is transparency, accountability, and liability. Without these three steps, detection for the people who need it most, provenance that is rights respecting, and that pipeline of responsibility, we're going to get stuck, looking in vain for the six fingered hand or the eyes that don't blink. We need to take these steps, otherwise we risk a world where it gets easier and easier to both fake reality and dismiss reality as potentially faked.

And that is a world that the political philosopher Hannah Arendt described in these terms: "a people that no longer can [00:46:00] believe anything cannot make up its own mind. It is deprived not only of its capacity to act, but also of its capacity to think and to judge, and with such a people, you can then do what you please".

That's a world I know none of us want, and that I think we can prevent.

BONUS - Deepfaking Democracy: Why AI Threatens News And Global Elections In 2024 Part 2 - Forbes - Air Date 2-6-24

ALEXANDRA S. LEVINE, FORBES WRITER: One of the most interesting pieces of this, and troubling pieces of this, is that in many cases the deepfake news segments that we found were getting more views and more virality than actual news segments from those same outlets that were posted to their blue check verified social media accounts around the same time.

One example that we found was from *Face the Nation*. This YouTube and TikTok creator had a segment that was actually one of his more innocuous segments that was about a group of kids jumping in an elevator and the elevator crashes down and then they owe this building more than half a million dollars in damages.

NEWS ANCHOR: Over 560,000 dollars in damages liable after TikToker Krishna [00:47:00] Sahai destroys elevators.

ALEXANDRA S. LEVINE, FORBES WRITER: So, it's not the most threatening example but what was so fascinating about it was that it was viewed more than 300,000 times and it used again the *Face the Nation* logo and on the *Face the Nation's* social media account on TikTok, the post from the same day only garnered 7,000 views.

So, when fake news segments from a creator that uses the outlet's logo or anchors from that station is in fact getting more eyeballs than actual news clips from the actual outlet's blue check social media accounts, you can see how that could become extremely problematic and deter people from actually following what is considered real news.

KEVIN GOLDBERG: So, even deepfake technology is protected by the first amendment, lying is protected by the First Amendment. I could make false statements and I am not going to be punished unless that false statement carries some additional harm with it. Direct harm. Usually harm that is perpetrated against an individual. And [00:48:00] frankly, when you're in, you know, you're in a situation talking, making political statements, that's the strongest protection the First Amendment gives.

So a general statement of a political nature, which I think a lot of deepfakes we're seeing in this coming year will be, are actually protected, even when

they're lies, unless there is some direct harm that is inflicted upon an individual or even, you know, to some degree, a small segment of society. And what we're talking about here are things like defamation. You know, if I say something or if I use a deepfake in a way that makes a false statement about you, harms your reputation, that would be something that is now outside of the First Amendment, not protected by the First Amendment.

While we do have a collective media literacy problem in this country, where people don't know the difference necessarily between news and opinion, or even within news, the difference between a good source and a not so good source or an outright lying source that has an agenda of its own, we're getting better with that. I think people collectively, and this is anecdotal, people collectively are getting better [00:49:00] at identifying, you know, separating truth from falsity. It's harder when you bring in video, because they don't know the same tells that we're already being trained to look for in printed or online information.

So, there's a level of validity, a veracity to something that they see in video and they go, Oh, it's video. It's really hard to fake that. And it's happening so much, and frankly, it's mostly being perpetrated by people who want to take advantage of it. You know, we know that in the 2020 and 2016 elections, a lot of the misinformation during the election period was coming from overseas. From places we aren't going to be able to get to, to, you know, to punish. And I think that's probably what's going to happen again, which is what makes it so difficult to combat.

ALEX MARQUARDT, CNN ANCHOR: This is the most comprehensive report that we've gotten about the 2020 election and foreign interference by the intelligence community and it does make clear that this massive Russian influence campaign was designed, orchestrated by Putin, to denigrate Joe Biden and to support the re [00:50:00] election of President Donald Trump.

HANY FARID: You know, I've heard people say, Look, disinformation, deepfakes, they can't change an election. And I don't think that's true, because if you look at the last two election cycles, the difference between one candidate or another in terms of the electoral vote came down to some 80,000 votes in a handful of states.

You don't have to move tens of millions of votes. You have to move tens of thousands of votes. And not only that, I know where those votes are. If I'm the bad guy trying to interfere with your election, I know exactly what states, I know exactly what towns, what localities, and I know how to find these people on social media and manipulate them.

That, I think, should worry us. You need a series of defenses, and so you need a series of proactive defenses and a series of reactive defenses, and you need better corporate responsibility, and you need some liability, and you need some regulation, and you need good consumer protection. And so, you know, when you put all those pieces together, I think we can start to trust things that we see online a little bit.

KEVIN GOLDBERG: It's possible. It's difficult. Both in a legal sense and a [00:51:00] practical sense to bring a defamation lawsuit against someone based on their creation of a deepfake. So let's just say you create a deepfake about me. I have to show very specifically that not only you lied, you harmed me in some way and specifically you harmed my reputation. But beyond that, I have to show a number of other things. I have to show a statement, specifically, that you made a materially and substantially false assertion of fact about me that was published and harmed my reputation and that you did it with some level of fault.

TIM BOUCHER, AI ANALYST: This ability to create so many images so rapidly, it's an incredibly powerful tool.

DONIE O'SULLIVAN, CNN CORRESPONDENT: New artificial intelligence technology makes it easy to create fake images that can look very realistic. Like these created by artist and online trust and safety expert, Tim Boucher.

HANY FARID: I think absolutely we are going to see the campaigns use it against their opponents. We also can see campaigns using it to bolster their own opponent, to create images of them looking more [00:52:00] heroic, or taller, for example. But here's the other place that we can, that the candidates can use it. Imagine now there's a hot mic of a candidate or a sitting president saying something inappropriate or illegal. They don't have to cop to it anymore. They can say it's fake. And so they can also deny reality. So, the deepfake technology is a double edged sword. You can create harmful content, but you can also dismiss real content by simply saying it's fake and muddying the waters.

ALEXANDRA S. LEVINE, FORBES WRITER: I think the most important thing right now is to remind people to think before they share, to be a bit skeptical of what they are consuming, and to really try to pay attention to the source. If the source is an authoritative news outlet, great. I don't think we can rely anymore on which accounts are blue check verified accounts and which aren't because now we know that many of the social media platforms allow people, any person, to purchase verification where the bar is significantly lower for verified accounts.

But I think that you should really be focused [00:53:00] on where the news is coming from, who is posting it, what their motive may be, and what sorts of perspectives they are including in the clip. I think all of those things are able to help us, especially in a very fast moving news environment, better calculate what is worth sharing versus what isn't, and help us better understand what we are consuming.

BONUS - The Taylor Swift Deepfake Saga - What Next: TBD | Tech, power, and the future - Air Date 2-2-24

EMILY PECK - HOST, WHAT NEXT: Tell me about the Telegram channel where these images were originating. What is it for? Who's in it? What do they talk about?

EMANUEL MAILBERG: It's like, imagine the id of a horny teenager. That's kind of the vibe. It's kind of dark. It's not a pleasant place to be, I have to be honest. It's a channel with tens of thousands of people. I don't want to be too specific so as not to direct people to it.

EMILY PECK - HOST, WHAT NEXT: Yeah.

EMANUEL MAILBERG: There are like sub communities within it. So, some of them are doing this deepfake stuff, some are doing photoshops, some are doing stuff... I'm just going to say it, and you can cut it out if you want, but it's like there's a tribute channel, right? And what is a tribute channel? A tribute channel is people [00:54:00] share photos of celebrities or people that they know in real life, and they film themselves masturbating against the images. And that's something that people enjoy doing. It's an underbelly of sexuality and online pornography that is not the kind of stuff that you would easily find on, say, a pornhub, but is readily available if you're still inclined. And many people are.

EMILY PECK - HOST, WHAT NEXT: And before these images were on Telegram, they were on 4chan. I mean, what does that say about the scope of this issue or problem, do you think?

EMANUEL MAILBERG: 4chan has been trying to find these loopholes and these free AI tools since they became available. So, last year, in November I think, we reported on this image of SpongeBob SquarePants doing 9/11. I don't know if you saw that.

EMILY PECK - HOST, WHAT NEXT: Should we be clear that SpongeBob didn't do 9/11?

EMANUEL MAILBERG: As far as we know, cannot confirm that he was [00:55:00] involved. And they did it with Bing, right? And Bing obviously does not want people to make images of 9/11 with their software. So you couldn't type in 'twin towers collapsing' or anything like that. But if you were to type in SpongeBob SquarePants in a cockpit of a jet flying towards two tall skyscrapers, it would generate the image, and it would look exactly like the twin Towers. So that is the kind of thing that they've been doing for months. And I think just recently it has become apparent that they found loopholes that allow them to do pornography.

EMILY PECK - HOST, WHAT NEXT: But, I mean, for all the horrors of Telegram and 4can, most people saw these images on X for the first time. Do you know how long they were there before everything kind of got amped up and went viral, et cetera?

EMANUEL MAILBERG: I think they went viral within 24 hours.

EMILY PECK - HOST, WHAT NEXT: Wow.

EMANUEL MAILBERG: And I would say within, I don't [00:56:00] know, 12 hours, 6 hours, the Swifties were on it, and they were pushing it down the feed, and by the next day, it was gone. Even X under Elon Musk with all the terrible content, I think this got so much heat that, I was surprised that it was removed that quickly, given the stuff that they do allow and the stuff that Musk himself puts out there.

EMILY PECK - HOST, WHAT NEXT: So it takes Taylor Swift to get Elon Musk's X to do any content moderation.

EMANUEL MAILBERG: It takes the biggest celebrity in the world, with the biggest, most devoted following in the world, to get Musk to move. Yeah, and the White House as well, right?

EMILY PECK - HOST, WHAT NEXT: I mean, what do regular people do? What do C-list celebrities do when something like this happens? What can they do?

EMANUEL MAILBERG: I mean, unfortunately, I hate to say this, but if you're not Taylor Swift, you're kind of screwed. And I see this all the time, and

it's heartbreaking and it's horrible. This is true both of minor celebrities, Instagram influencers, Twitch streamers, YouTubers [00:57:00] who are deepfaked regularly. It happens every day. I see it every day in my reporting. And they either don't know that it's happening, and if they do know that it's happening and you approach X or you approach whatever platform that is hosting and enabling that content, wherever it is hosted, chances are they'll do something about it. But that puts those people in the impossible position of policing the entire Internet to remove that content, and it's not possible. We know people who have tried to do this, and it's not only very hard to do, it's retraumatizing. We see this with what's colloquially called revenge porn. It's a terrible process. Some people try, some people, it's like too painful for them to even pursue it. There's no good answer. Part of the amazing thing about the Taylor Swift story is that you do see action. You see action from Microsoft, you see action from X, you see policy efforts, and [00:58:00] you're not going to get this as a normal person or a minor celebrity.

EMILY PECK - HOST, WHAT NEXT: So, I mean, as you've explained, deepfakes took some technical know-how and some effort, but now these image generators, I mean, they're really easy to use. I even tried the Microsoft tool earlier today and my prompts were generic and boring. So, I'm not a visual artist in any way, but I mean, it's really easy to use these things. Are we just facing down a potential, just, explosion of these kinds of images of most people? I mean, most women and girls?

EMANUEL MAILBERG: Yeah, we're in it. It's very important to make clear that this is primarily targeting women. Overwhelmingly targeting women. And there's data to back this up. People often talk about the political implications of deepfake and misinformation. And when you look at the data, that is not what happens. Most of what people are doing with it is creating non-consensual images of women. We're in the thick of it. Like, it's happening. The [00:59:00] explosion is here, we're in the middle of it. The good news is that I truly don't think that it's going to stay this way. Because when we report on this stuff, the companies that make these tools are embarrassed and horrified, and they make changes. And the thing that we're doing right now is going case by case, company by company, image by image, and reporting on it. And we're seeing results. Like, improvements are being made. But I think that in order to see a big improvement, I think something worse is going to have to happen. I think we're going to have to see some truly horrible, either it's a specific case that goes to court and somebody gets sued, or it's like some viral media story about somebody who got really hurt. We need to hit rock bottom, I think, in a way, before we really see big changes.

EMILY PECK - HOST, WHAT NEXT: Emanuel sees a precedent here: another time when there were big changes in porn on the Internet. And that's what happened at Pornhub.

EMANUEL MAILBERG: When we started reporting on [01:00:00] Pornhub, the state of the platform was that anyone can upload any video. Obviously, because that was the case, we were reporting on many cases of abuse, and we spent a few years reporting on this, reaching out to Pornhub for comment and telling them what we're seeing and publishing stories about it. And they were very dismissive. It was always like, You know, we have, it's like we have moderation methods and you can issue takedown requests and we're responsive and responsible and blah, blah, blah. But the abuse continued and then the lawsuit started to pile up. There was child abuse. There's this big GirlsDoPorn case where 400 women were exploited by this porn company that published its videos on Pornhub. And it got to a point where the platform really had to change. They purged it of millions of videos and they changed the rules of the platform, where now every single person who is in a video on Pornhub needs to provide written, active consent for them to appear in the video. Pornhub changed its name, it changed its ownership. [01:01:00] It's a completely different Internet platform, but it only became that way because things got really bad. And I think that's the path we're on.

Final comments on even more dangers from news sites populated with AI-generated content

JAY TOMLINSON - HOST, BEST OF THE LEFT: We've just heard clips today, starting with *Forbes*, looking at reactive versus proactive approaches to detecting manipulated media. *CYBER* advocated for new federal legislation to regulate tech firms. *All Things Considered* looked at big tech taking baby steps on self-regulation. *Aperture* discuss the real world harms that AI is already having on women and girls around the world. *On with Kara Swisher* considered the difficult balance of freedom of speech and online regulation. And *TED Talks Daily* looked at three tools to create an infrastructure of authenticity.

That's what everybody heard, but members also heard bonus clips from *Forbes* looking further into deepfake news stories made to look like legitimate mainstream outlets, and *What Next: TBD* analyzed the impact of the recent Taylor Swift deepfakes [01:02:00] event. To hear that and have all of our bonus contents delivered seamlessly to the new members only podcast feed that you'll receive, sign up to support the show at bestoftheleft.com/support, or shoot me

an email requesting a financial hardship membership, because we don't let a lack of funds stand in the way of hearing more information.

Now to wrap up, I just want to add one more thing, at least. I know there's more than one, but for now, I'll add one more thing that you should be worried about regarding AI. Which is that the widespread availability of AI text generation has begun to make it much, much more likely that a random new site you may stumble upon through just a regular Google search will actually just be a wasteland of clickbait content designed to rank highly in search results, but give no valuable information. Sometimes these sites will even manage to get the URLs of former legitimate news sites.

So, for instance, two local newspapers might merge [01:03:00] and one or both of the original URLs stops being the official site of the new merged paper. Then, if those old URLs are allowed to lapse, you know, the company doesn't keep up on payments and they're allowed to go back on the market, one of these clickbait farms can and likely will grab it and repopulate that site with their own AI generated content, making it look more legitimate thanks to the seemingly trustworthy URL, maybe even a URL that users had been used to going to and trusting for years. So, it's becoming even more important now to get your news from trusted sites or trusted aggregators. For instance, Apple News or Google News or apps like a Ground News, won't be likely to feature articles from unverified sites. You know, there may be a case where they'll get tricked into it, but generally that won't be the case just as, you know, we certainly do [01:04:00] our due diligence researching any news sources that we take on before we feature them here on the show.

But as it was hopefully made clear in the show today, this is yet another systemic problem requiring systemic solutions. Humanity will have very little chance of overcoming the problem of AI disinformation with a simple libertarian, buyer-beware sort of approach. We are just not wired to function in a world where we have to disbelieve everything we see and hear until it can be proven beyond a reasonable doubt. It's actually an evolutionary benefit of ours to have a basic tendency towards trust in other people and in what we see. Human superpower is our ability to work together in flexible ways and that requires trust. We couldn't have evolved to be able to build the complicated society we have today, that can actually sustain the number of humans on the planet, [01:05:00] without bad tendency towards trust. We would have gotten stuck way back along the evolutionary line.

And to be sure that tendency to trust has been exploited by bad actors all throughout human history, but that tendency towards trust has maintained itself. But now, the ability of bad actors to use people's trusting nature against them is

reaching a truly unprecedented level. And we really need to understand it as the existential threat that it is, or existential threat to democracy at least. Society will probably be able to carry on either way, but it may just be in the form of autocratic rule over a subjugated people because as has been well known since at least Thomas Jefferson's time, that, as he said, "an educated citizenry is a vital requisite for our survival as a free people". And that's simply not something that is possible to have when [01:06:00] people are a wash in disinformation.

That is going to be it for today. As always keep the comments coming in. I would love to hear your thoughts or questions about this or anything else. You can leave a voicemail or send us a text at 202-999-3991, or simply email me to jay@bestoftheleft.com. Thanks to everyone for listening. Thanks to Deon Clark and Erin Clayton for their research work for the show and participation in our bonus episodes. Thanks to our Transcriptionist Trio, Ken, Brian, and Ben, for their volunteer work helping put our transcripts together. Thanks to Amanda Hoffman for all of her work on our social media outlets, activism segments, graphic designing, web mastering, and bonus show co-hosting. And thanks to those who already support the show by becoming a member or purchasing gift memberships. You can join them by signing up at bestoftheleft.com/support, through our Patrion page, or from right inside the Apple podcast app. Membership is how you get instant access to our incredibly good and often funny bonus episodes, in addition to there [01:07:00] being extra content, no ads, and chapter markers in all of our regular episodes, all through your regular podcast player. You'll find that link in the show notes, along with a link to join our Discord community where you can also continue the discussion.

So, coming to from far outside, the conventional wisdom of Washington DC, my name is Jay, and this has been the *Best of the Left* podcast, coming to twice weekly thanks entirely to the members and donors to the show, from bestoftheleft.com.