AUTHENTICATED
U.S. GOVERNMENT
INFORMATION
GPO

S. HRG. 118–738

PROTECTING CONSUMERS FROM ARTIFICIAL INTELLIGENCE ENABLED FRAUD AND SCAMS

HEARING

BEFORE THE

SUBCOMMITTEE ON CONSUMER PROTECTION, PRODUCT SAFETY, AND DATA SECURITY

OF THE

COMMITTEE ON COMMERCE, SCIENCE, AND TRANSPORTATION
UNITED STATES SENATE

ONE HUNDRED EIGHTEENTH CONGRESS

SECOND SESSION

NOVEMBER 19, 2024

Printed for the use of the Committee on Commerce, Science, and Transportation

Available online: http://www.govinfo.gov

U.S. GOVERNMENT PUBLISHING OFFICE
61–951 PDF WASHINGTON : 2025

SENATE COMMITTEE ON COMMERCE, SCIENCE, AND TRANSPORTATION

ONE HUNDRED EIGHTEENTH CONGRESS

SECOND SESSION

MARIA CANTWELL, Washington, *Chair*

AMY KLOBUCHAR, Minnesota
BRIAN SCHATZ, Hawaii
EDWARD MARKEY, Massachusetts
GARY PETERS, Michigan
TAMMY BALDWIN, Wisconsin
TAMMY DUCKWORTH, Illinois
JON TESTER, Montana
KYRSTEN SINEMA, Arizona
JACKY ROSEN, Nevada
BEN RAY LUJÁN, New Mexico
JOHN HICKENLOOPER, Colorado
RAPHAEL WARNOCK, Georgia
PETER WELCH, Vermont

TED CRUZ, Texas, *Ranking*
JOHN THUNE, South Dakota
ROGER WICKER, Mississippi
DEB FISCHER, Nebraska
JERRY MORAN, Kansas
DAN SULLIVAN, Alaska
MARSHA BLACKBURN, Tennessee
TODD YOUNG, Indiana
TED BUDD, North Carolina
ERIC SCHMITT, Missouri
J. D. VANCE, Ohio
SHELLEY MOORE CAPITO, West Virginia
CYNTHIA LUMMIS, Wyoming

LILA HARPER HELMS, *Staff Director*
MELISSA PORTER, *Deputy Staff Director*
JONATHAN HALE, *General Counsel*
BRAD GRANTZ, *Republican Staff Director*
NICOLE CHRISTUS, *Republican Deputy Staff Director*
LIAM MCKENNA, *General Counsel*

SUBCOMMITTEE ON CONSUMER PROTECTION, PRODUCT SAFETY, AND DATA SECURITY

JOHN HICKENLOOPER, Colorado, *Chair*
AMY KLOBUCHAR, Minnesota
BRIAN SCHATZ, Hawaii
EDWARD MARKEY, Massachusetts
TAMMY BALDWIN, Wisconsin
TAMMY DUCKWORTH, Illinois
BEN RAY LUJÁN, New Mexico
PETER WELCH, Vermont

MARSHA BLACKBURN, Tennessee, *Ranking*
DEB FISCHER, Nebraska
JERRY MORAN, Kansas
DAN SULLIVAN, Alaska
TODD YOUNG, Indiana
TED BUDD, North Carolina
CYNTHIA LUMMIS, Wyoming

CONTENTS

WITNESSES

APPENDIX

PROTECTING CONSUMERS FROM ARTIFICIAL INTELLIGENCE ENABLED FRAUD AND SCAMS

TUESDAY, NOVEMBER 19, 2024

U.S. SENATE,
SUBCOMMITTEE ON CONSUMER PROTECTION, PRODUCT
SAFETY, AND DATA SECURITY,
COMMITTEE ON COMMERCE, SCIENCE, AND TRANSPORTATION,
Washington, DC.

The Subcommittee met, pursuant to notice, at 2:39 p.m., in room SR–253, Russell Senate Office Building, Hon. John Hickenlooper, Chairman of the Subcommittee, presiding.

Present: Senators Hickenlooper [presiding], Klobuchar, Schatz, Baldwin, Luján, Markey, Blackburn, and Sullivan.

OPENING STATEMENT OF HON. JOHN HICKENLOOPER, U.S. SENATOR FROM COLORADO

Senator HICKENLOOPER. Now, this meeting is called to order.

Welcome to the Subcommittee on Consumer Protection, Product Safety, and Data Security. We are now in order.

We stand today—I think everyone agrees that we are at a crossroads where American leadership in AI is going to depend on which of many courses Congress takes, going forward.

This has been, can, and should remain a nonpartisan or let us say a bipartisan effort that focuses on certain core issues: promoting transparency in how developers build new models, adopting evidence-based standards to deliver solutions to problems that we are aware of that we know of that exist in AI, and third, building Americans' trust in what could be a and has been on occasion a very disruptive technology.

I think our constituents in our home states but across the country and really around the world are waiting for us to take the reins and strengthen America's leadership but at the same time not compromising our commitment to innovation and transparency in AI.

I believe in my heart and soul—I think most of us on this committee believe that American leadership is essential for AI for a whole variety of reasons. I think we will go through a lot today.

We already know that the capabilities in the field of artificial intelligence, those capabilities are evolving, changing rapidly.

Generative AI tools allow almost anyone to create realistic, synthetic video—well, synthetic text, image, audio, video, you name it. Whatever content you can imagine we can now create.

And while AI will have enormous benefits—and I truly believe that—benefits in our daily lives in sectors like clean energy, medicine, workplace productivity, workplace safety—for all those benefits we have to mitigate and anticipate the concurrent risks that this technology brings along with it.

Just one example, this image behind me. Is there a poster—oh, there we are. This was created using AI tools depicting a young girl holding a dog in the aftermath of Hurricane Helene. We had a label to clearly show that the image is AI generated.

While not a real image, it appears to be extremely realistic at first glance, although I grilled my staff on exactly what were the details—how could they, the trained observer, recognize this as synthetic.

But I think without a clear label it really does take experience and training, a much closer look, to see the flaws in the image—the young girl's left hand somewhat misshapen—from a natural photograph. I will not go into it. I was so proud of myself I could follow the argument.

But I think we recognize and should all own up to the fact that scammers are already using this new technology to prey on innocent consumers.

There are a number of bad actors out there that see this as a vehicle for sudden and dramatic enrichment. One example—scammers have cloned the voices of loved ones saying they have been kidnapped, they have been abducted, and the—this familiar voice is begging for ransom payments.

Other deepfake videos show celebrities endorsing products or candidates who they had never endorsed and really had no intention of endorsing.

Many troubling examples exhibit that include children, teens, women depicted in nonconsensual intimate—nonconsensual, intimate and violent images or videos that in many occasions cause deep emotional harm.

These AI-enabled scams do not just cost consumers financially, but they damage reputations and relationships and, equally important, they cast doubt about what we see and what we hear online.

As AI-generated content gets more elaborate, more realistic, almost anyone can fall for one of these fakes. I think we have to—we have a responsibility to raise the alarm for these scams and these frauds and begin to be a little more aggressive in what can be done to avoid them.

During our hearing today we will, you know, begin to understand and look at some of the tools and techniques companies and consumers can use to recognize malicious deepfakes to be able to discuss which public and private efforts are needed to educate the public with the experiences and the skills necessary to avoid AI-enabled scams, and then, third, to highlight enforcement authorities that we can establish to deter bad actors and prevent further harm coming to consumers.

This committee is already at work on this, has already drafted, amended, passed several bipartisan bills focused on AI issues. I will just run through these.

The Future of AI Innovation Act fosters partnerships between government, private sector, and academia to promote AI innovation. Validation and Evaluation for Trustworthy AI Act creates a voluntary framework which will enable third party audits of AI systems.

AI Research, Innovation, and Accountability Act increases R&D into content authenticity, requires consumer AI transparency, and creates a framework to hold AI developers accountable.

And then, last, the COPIED Act—C-O-P-I-E-D—COPIED Act has not yet been considered by the Committee but increases Federal R&D in synthetic contact detection and creates enforceable rules to prevent bad actors from manipulating labels on content.

Last, the Take It Down Act makes it a criminal offense to create or distribute nonconsensual intimate images—NCII—of individuals.

These are each bipartisan bills. They lay the groundwork for responsible innovation to address real problems with thoughtful solutions. They are not perfect and I trust will come out of information where we will get improvements from your sage wisdom.

We look forward to working hard together to get these across the finish line and passed into law in the coming weeks, but we know that the bad actors, unfortunately, still continue to try and use this technology for fraudulent purposes.

To combat fraud the Federal Trade Commission—the FTC—recently adopted rules to prohibit the impersonation of government agencies and businesses including through the use of AI.

The FTC is also considering extending this protection to individuals including through visual or audio—visual or audio deepfakes. This is one very good example of a Federal agency taking decisive action to address a specific harm. But we need to all encourage further targeted, specific efforts with this basic common sense rule.

States across the country have begun to enact legislation—states being the laboratories of democracy—to try and address the creation and distribution of deepfake media. Again, the map behind me here shows states taking action.

The yellow states have enacted legislation related to AI use in election contexts. States in purple have enacted legislation related to nonconsensual intimate imagery, and states in red have enacted legislation related to both of these and instead—or instead to address other AI-generated media concerns. [Map is shown]

And as we can see this is right now a patchwork of protections which is defying the need for predictability, which pretty much any industry needs to prosper.

A consistent Federal approach would be tremendously beneficial to fill in these gaps and make sure we are protecting all Americans.

Promoting responsible AI also rely on our partners in the private sector. A number of companies—Anthropic, Google, Microsoft, OpenAI, a number of others—have made voluntary commitments to responsible AI practices. These include commitments to help Americans understand whether types of content they see is AI generated. It could also be done through watermarks or similar technology that identifies the origin, ownership, or permitted uses of a piece of AI-generated content.

Today, we are going to hear from leading experts, all of you, in artificial intelligence and AI-generated media about what—from your perception what is already happening but more importantly what else needs to be done, where we should be focusing.

Hopefully, working together we can do a better job of protecting Americans from these potential risks and the scams and frauds but at the same time make sure that we unleash the innovation that this country is known for, especially in this emerging technology.

I do not have to go into the importance of the competition around AI, that this is a competition that some of our global rivals take very seriously, and if we are any less focused on that that will be to our own detriment.

I want to welcome each of our witnesses who are joining us today. Dr. Hany Farid, Professor of the University of California at Berkeley School of Information; Justin Brookman, Director of Privacy and Technology Policy at Consumer Reports; Mr. Mounir Ibrahim, Chief Communications Officer and Head of Public Affairs at Truepic; and Ms. Dorota Mani—you guys are running me ragged. With a name like Hickenlooper I usually get granted a certain latitude. Ms. Mani is an entrepreneur, mother of a victim of non-consensual intimate imagery.

I would now like to recognize Ranking Member Blackburn for her opening remarks.

STATEMENT OF MARSHA BLACKBURN, U.S. SENATOR FROM TENNESSEE

Senator BLACKBURN. Thank you, Mr. Chairman.

I am absolutely delighted that we are having this hearing today and focusing on scams and fraud.

Now, in Tennessee I say we have got the good, bad, and ugly relationship with AI. A lot of our manufacturers, people that are working in healthcare, predictive diagnosis, disease analysis, logistics, they are utilizing AI every day to achieve efficiencies.

Our songwriters and our entertainers are saying, hey, wait a minute, we got to have some help with this, and you reference the COPIED Act, the No Fakes Act that some of my colleagues and I have done on a bipartisan basis.

And then the ugly is really what is happening to people with these scams and with these frauds, and especially when it comes to senior citizens and what we are seeing happen there.

Now, I thought it was very interesting that the FTC with the Consumer Sentinel Network Data Book they listed that scams increased a billion dollars over the last 12 months to $10 billion.

This is what it is costing, and from a year prior it was up a billion dollars. When you think about the rise in this you have to look at what this is doing and, of course, we know AI is what is driving a lot of this.

It is a technology that is advancing so incredibly fast and, of course, legislation never keeps pace with technology.

So we know that it is catching a lot of people that are really quite savvy when it comes to conducting a lot of their transactional life online and we know that older Americans are the ones who have been hit most frequently with this as an emerging threat.

Now, the fraudsters that are utilizing AI to deceive consumers are—have gotten crafty when it comes to creating these highly personalized and convincing attacks and I think what is surprising to a lot of people is the fact that they are doing this at scale, and the replication of these attacks and the tailored e-mails, the text messages, the images—as the Chairman showed the altered images—and those are used to trick people to click that link, and then once they have clicked that link they are then downloading malware.

They are divulging personal information and the fraudsters feel like that they have got them. But they have become—these attacks have become very realistic. The spear phishing e-mails that really use a lot of this make it appear that it is coming from a trusted source and the—adding to this the chat box, which make it appear that you are having an actual real-time conversation with someone, is very disarming.

So we know that these are becoming—the use of these tools by the scammers are becoming more prevalent. They are becoming more precise, more widespread, and harder to attack.

And when we get to what do you do about this and how do you combat this we know that it does require an encompassing approach.

It has got to be comprehensive and, of course, consumers need to be armed with knowledge about what is happening here and also looking to improve their digital hygiene and their digital literacy so that they know more about what they need to look for with those red flags.

And we know that it is also going to require that we move forward with having an actual online privacy standard, which we have never passed.

In this committee and when I was in the House, we continued to look at this so that individuals have the ability to actually protect their virtual you, which is their presence in the virtual space, and it is going to require that we take those actions.

We are really thrilled to have you all before us today. It helps to inform not only our Committee, our Subcommittee, informs our colleagues and builds into the record for the need to move forward on legislation that will enhance the privacy and help to protect our citizens in the virtual space.

So welcome to each of you. We look forward to hearing your statements.

Senator HICKENLOOPER. Thank you, Senator Blackburn.

Now we will ask for each of your statements, one after the other.

First, let me introduce—reintroduce again Dr. Hany Farid, Professor of—at the University of California Berkeley School of Information. He is also a member of the Berkeley Artificial Intelligence Lab, Senior Faculty Advisor for the Center for Long-Term Cybersecurity.

I have a son who is finishing. He is a senior doing engineering and computer science at Stanford and he tells me that Berkeley is an inferior institution.

[Laughter.]

Senator HICKENLOOPER. I have heard there is a lot of contention there.

Dr. Farid.

STATEMENT OF HANY FARID, PH.D., PROFESSOR, UNIVERSITY OF CALIFORNIA BERKELEY, ELECTRICAL ENGINEERING AND COMPUTER SCIENCE, SCHOOL OF INFORMATION

Mr. FARID. I will not hold it against you, Senator.

Thank you and thanks for having me. I am by training an applied mathematician and computer scientist, and for the past 25 years as an academic, I have been working to create and deploy technologies that combat the spread of child sexual abuse material, combat online extremism and dangerous disinformation campaigns, and most recently combating the ugly world of deepfake imagery.

I would like to start with a quick overview of today's world of AI in which there are two main branches, predictive AI and generative AI, and when I know we will talk a lot about generative AI we should look at the entire ecosystem here because I think there are issues on both sides.

Predictive AI is tasked with predicting, anything from the movies you want to watch on Netflix to who will default on a loan, who will be a high-performing employee, who will recidivate if given bail, to what neighborhood should be patrolled by law enforcement.

Generative AI is tasked with creating content from texts to images to audio to video. For text this can mean answering simple questions, summarizing a brief, helping a student with their math homework.

But there are also troubling applications where we have seen interactive AI bots give instructions on how to create dangerous weaponry and encouraging someone to harm themselves or someone else.

For images, audio, and video while there are also many creative applications we are seeing some deeply troubling applications, from the creation of nonconsensual intimate imagery, child sexual abuse, to imposters in audio and video used to perpetrate small- to large-scale fraud, everything you enumerated in your opening remarks.

Although not fundamentally new, new advances in machine learning have over the past few years fueled rapid developments in both of these areas with, not surprisingly, very exciting applications but also some worrisome applications.

So as we consider how to embrace and, yet, protect ourselves from this latest wave of technology I would like to offer a few thoughts.

When it comes to predictive AI it is useful to consider a risk or harms-based approach. Predictive algorithms that make movie recommendations should be considered in a different category from algorithms that make decisions regarding our civil liberties, our finances, our employment.

For the latter, we have seen troubling failures of so-called black box algorithms that are not well understood or auditable. Most notably, predictive algorithms being used today in policing and criminal justice have been found to be biased against people of color and algorithms being used in H.R. departments have been found to be biased against women.

Fundamentally, the limitation of today's predictive AI is that they are neither artificial nor are they intelligent.

They are pattern matching solutions that look at historical data and recreate historical patterns. So if the historical data is biased then your AI-powered future will be equally biased.

Today's predictive AI is not much more than a sophisticated parrot. When it comes to generative AI the harms I enumerated earlier were completely predictable. Before the more polite term generative AI was coined we referred to this content as deepfakes which has its very roots in the moniker of a Reddit user who used the then nascent AI technology to create the earliest examples of nonconsensual intimate imagery.

The harms we are seeing today from deepfakes are neither unintended nor unexpected. They are baked into the very DNA of this technology, and as Senator Blackburn mentioned in her opening remarks, the theft of the intellectual property that is driving all of this should also worry everybody greatly.

So when considering mitigation strategies we need to consider not only the underlying AI technology but the entire technology ecosystem. So with respect to NCII and CSAM, for example, there are many players that fuel the harms.

So there is, of course, the person creating the content. There is the tool used to create the content. But then there is the service used to distribute the content, mostly social media, but also the financial institutions that enable the monetization of the content—web services from search engines that gleefully surface this content to cloud computing and domain name services that provide the infrastructure for bad actors.

Importantly, because AI is not the first nor will it be the last example of technology weaponized, we should not become overly focused on just AI but we have to seek broader thinking about the technology when we are looking at mitigation strategies.

Last, you will hear many loud voices—many of them from Stanford, by the way—that claim that reasonable guardrails to make products and technology safer will destroy innovation.

I reject these claims as nothing more than self-serving and blind or indifferent to the actual harms that are facing individuals, organizations, and societies. We fell for these same lies, by the way, from the same voices for the past 25 years and failed to rein in the worst abuses of social media.

We need not make those same mistakes in this latest AI revolution.

Thank you.

[The prepared statement of Mr. Farid follows:]

PREPARED STATEMENT OF HANY FARID, PH.D., PROFESSOR,
UNIVERSITY OF CALIFORNIA BERKELEY, ELECTRICAL ENGINEERING AND COMPUTER SCIENCE, SCHOOL OF INFORMATION

Biography

Hany Farid is a Professor at the University of California, Berkeley with a joint appointment in Electrical Engineering & Computer Science and the School of Information. His research focuses on digital forensics, image analysis, and human perception. He received his undergraduate degree in Computer Science and Applied Mathematics from the University of Rochester in 1989, and his Ph.D. in Computer Science from the University of Pennsylvania in 1997. Following a two-year post-doctoral fellowship in Brain and Cognitive Sciences at MIT, he joined the faculty at Dartmouth College in 1999 where he remained until 2019. He is the recipient of an Alfred P. Sloan Fellowship, a John Simon Guggenheim Fellowship, and is a Fellow of the National Academy of Inventors.

Testimony

Overview

Synthetic media—so-called deepfakes—have captured the imagination of some and struck fear in others. Although they vary in their form and creation, deep-fakes refer to text, image, audio, or video that has been automatically synthesized by a

AI-powered system. While not fundamentally new, today's enhanced ability to easily create, distribute, and amplify manipulated media comes with heightened risks. Reasonable and proportional interventions can and should be adopted that would allow for the creative uses of these powerful new technologies while mitigating the risk they pose to individuals, organizations, and democracies.

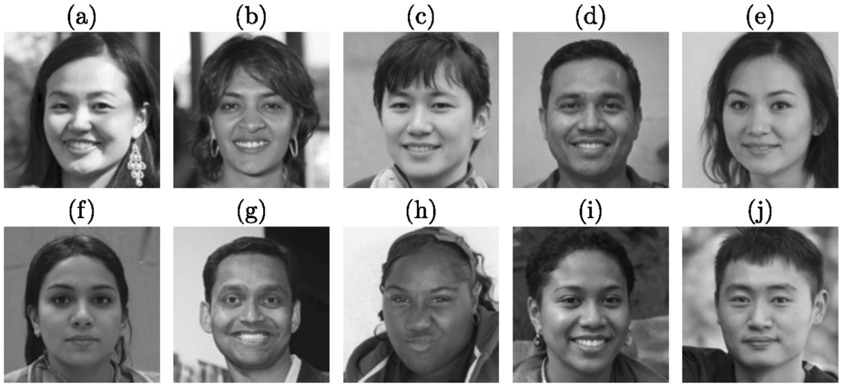


Figure 1: Half of these faces are real and half are AI generated. Can you tell which is which? See footnote at bottom of page for correct answers.

Deepfakes: Creation

Image

A generative adversarial network (GAN) is a common computational technique for synthesizing images of people, cats, planes, or any other category: generative because these systems are tasked with generating an image; adversarial because these systems pit two separate components (a generator and a discriminator) against each other; and network, because the computational machinery underlying the generator and discriminator are deep neural networks (hence the term deepfake).

StyleGAN is one of the earliest and most successful systems for generating realistic human faces, Figure 1. When tasked with generating a face, the generator starts by laying down a random array of pixels and feeding this first guess to the discriminator. If the discriminator, equipped with a large database of real faces, can distinguish the generated image from real faces, the discriminator provides this feedback to the generator. The generator then updates its initial guess and feeds this update to the discriminator in a second round. This process continues with the generator and discriminator adversarially competing until an equilibrium is reached when the generator produces an image that the discriminator cannot distinguish from real faces.[1]

Although highly realistic, GANs generally do not afford much control over the appearance or surroundings of the synthesized face. By comparison, more recent text-to-image (or diffusion-based) synthesis affords more rendering control. Trained on billions of images with an accompanying descriptive caption, each training image is progressively corrupted until only visual noise remains. The model then learns to denoise each image by reversing this corruption. This model can then be conditioned to generate an image that is semantically consistent with any text prompt like "an experienced word carver at work," Figure 2.

[1] The faces in panels Figure 1(a), (b), (g), (h), and (j) are real; the faces in panels (c), (d), (e), (f), and (i) are AI generated.

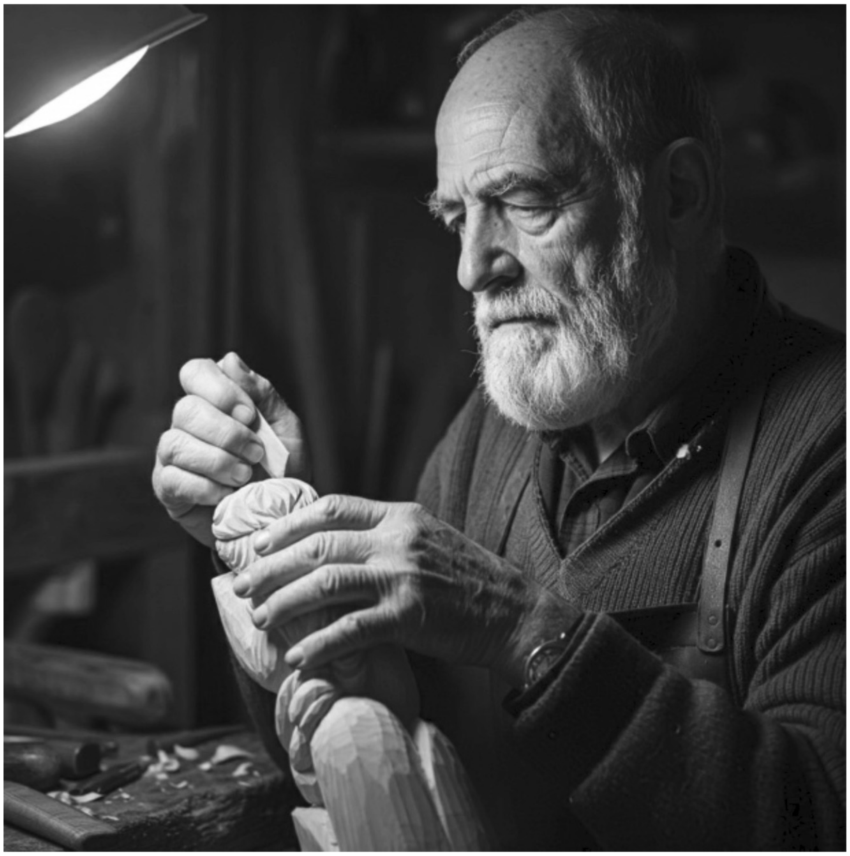


Figure 2: An AI-generated image created with the prompt "an experienced word carver at work."

Video

Video deepfakes fall into two broad categories: text-to-video and impersonation.

Text-to-video deepfakes are the natural extension of text-to-image where a model is trained to generate a video to be semantically consistent with a text prompt. A year ago, these systems tasked with creating short video clips from a text prompt like "Will Smith eating spaghetti" yielded videos of which nightmares are made.[2]

A typical video consists of 24 to 30 still images per second. Generating many realistic still images, however, is not enough to create a coherent video. These earlier systems struggled to create temporally coherent and physically plausible videos in which the inter-frame motion was convincing. Just a year later, however, these systems have improved tremendously. While not perfect, the resulting videos are stunning in their realism and temporal consistency, and quickly becoming difficult to distinguish from reality.

[2] *https://www.youtube.com/watch?v=XQr4Xklqzw8*

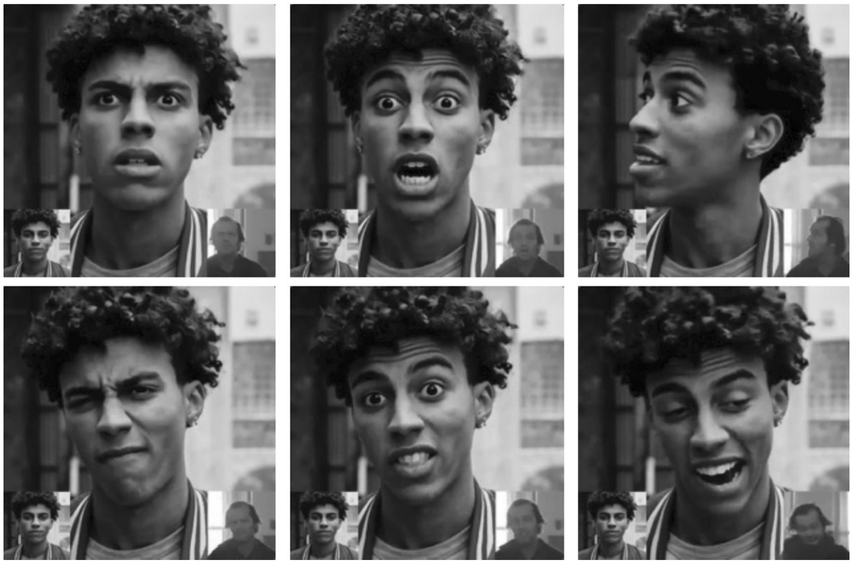


Figure 3: Selected frames of an avatar deepfake in which from a single photo of a person (bottom left), they are animated based on the movement of another person (bottom right).

Although there are several different incarnations of impersonation deepfakes, two of the most popular are lip-sync and face-swap deepfakes.

Given a source video of a person talking and a new audio track (either AI-generated or impersonated), a lip-sync deepfake generates a new video track in which the person's mouth is automatically modified to be consistent with the new audio track. And, because it is relatively easy to clone a person's voice from as little as 30 seconds of their voice, lip-sync deepfakes are a common tool used to co-opt the identity of celebrities or politicians to push various scams and disinformation campaigns.

A face-swap deepfake is a modified video in which one person's identity, from eyebrows to chin and cheek to cheek, is replaced with another identity. This type of deepfake is most common in the creation of non-consensual intimate imagery. Face-swap deepfakes can also be created in real time, meaning that you will soon not know for sure if the person at the other end of a video call is real or not.

And, the latest incarnation of impersonation deepfakes are puppet-master or avatar deepfakes in which a single image of a person is animated based on the movement and speech of another person, Figurefig:avatar.

The trend of the past few years has been that all forms of image, video, and audio deepfakes continue their ballistic trajectory in terms of realism, ease of use, and accessibility.

Deepfakes: Passing Through the Uncanny Valley

First coined by Japanese roboticist Masahiro Mori in the 1970s, the term uncanny valley describes a phenomenon that occurs when a humanoid robot, or an image or video of a computer-generated human, becomes more human-like. There is a point at which the humanoid depiction becomes eerily similar to humans but is still distinguishable from real humans, causing a significant drop in our emotional comfort and acceptance. This transition is known as the uncanny valley. A humanoid depiction is said to exit the uncanny valley when it becomes so realistic that it is indistinguishable from a real person. Generative AI is well on its way to passing through the uncanny valley.

Half of the faces in Figure 1 are real and half are AI generated. Can you tell which is which? If you are like most others, your performance on this task was at near chance. A recent perceptual study found that when asked to distinguish between a real and AI-generated face, participants performed no better than guessing. In a second study in which participants were provided with training prior to completing the task, their performance improved only slightly. AI-generated faces are highly realistic and extremely difficult to perceptually distinguish from reality.

Performance is only slightly better for videos of people talking. For AI-cloned voices, a recent study found that participants mistook the identity of an AI-generated voice for its real counterpart 80 percent of the time, and correctly identified a voice as AI-generated only 60 percent of the time.

While not all forms of AI-generated content have passed through the uncanny valley, what remains will almost certainly follow in the near future. We are quickly entering an era where it is increasingly more difficult for the average person to distinguish between fact and fiction.

Deepfakes: The Good

Hardly a day goes by when I don't use some form of generative AI in my work. From using large language models (LLMs) to write or debug code to using image synthesis to create visuals for a lecture. I cannot recall any other technology that has so dramatically and so quickly altered the way I work (and in some cases, think). My colleagues and students report a similar impact in their work and studies.

Beyond personal uses cases, a particularly empowering example of the use of generative AI was by Representative Wexton of Virginia who used an AI-generated version of her voice to address lawmakers on the House floor: "My battle with progressive supranuclear palsy, or PSP, has robbed me of my ability to use my full voice and move around in the ways that I used." Because today's generative AI can clone a person's voice from as little as a 30 second recording, Rep. Wexton was able to speak in her own voice as opposed to the tinny and slightly creepy computer-generated voices of just a few years ago.

I have little doubt that generative AI is and will continue to offer positive and exciting use cases, and be an intellectual and creative accelerant, but with a few caveats. All forms of generative AI have been trained on decades of user-generated content, in many cases without permission and in many cases in direct violation of copyright laws. Trying to justify their indiscriminate scraping of online content, OpenAI—one of the leaders in the generative-AI space—admitted it would be "impossible to train today's leading AI models without using copyrighted materials." This is a bit of a hard pill to swallow for a company that in less than 10 years has grown to a valuation of $150 billion.

Deepfakes: The Bad and The Ugly

Non-Consensual Intimate Imagery

Before the less-objectionable term "Generative-AI" took root, AI-generated content was referred to as "deepfakes", a term derived from the moniker of a Reddit user who in 2017 used this nascent technology to create non-consensual intimate imagery, NCII (often referred to by the misnomer "revenge porn," suggesting somehow that the women depicted inflicted a harm deserving of revenge). Seemingly unable to shake off; its roots, generative AI continues to be widely used to insert a person's likeness (primarily women and also children) into sexually explicit material which is then publicly shared by its creators as a form of humiliation or extortion.

While it used to take thousands of images of a person to digitally insert them into NCII, today only a single image is needed. This means that the threat of NCII has moved from the likes of Scarlett Johansson, with a large digital footprint, to anyone with a single photo of themselves online.

Shown in Figure 4, for example, is an image that I generated using a free service (that doesn't allow the generation of explicit material) in which I inserted my face into an image of an inmate in an orange jumpsuit.

A recent study surveyed 16,000 respondents in 10 countries and found that 2.2 percent of respondents reported being a victim of NCII, and 1.8 percent reported creating NCII. Given that many people may not know they are victims and many may be unwilling to admit creating NCII, this is surely a lower bound. The threats of NCII are not hypothetical nor are they relegated to the dark recesses of the internet. Finding NCII content and creation tools is no further than a Google search away.

Child Sexual Abuse Imagery

The Cyber Tipline at the U.S.-based National Center for Missing and Exploited Children (NCMEC) is a national reporting system for reporting all forms of child sexual exploitation including apparent child sexual abuse material (CSAM). The majority of reports come from electronic service provides including the largest social media platforms like Facebook and Instagram. In 2010, NCMEC received 132,000 reports. By 2015, the number of reports grew to over 4 million, and then 21 million in 2020 and 36 million in 2023. The average age of a child depicted in this content is 12 and sometimes as young as just a few months.

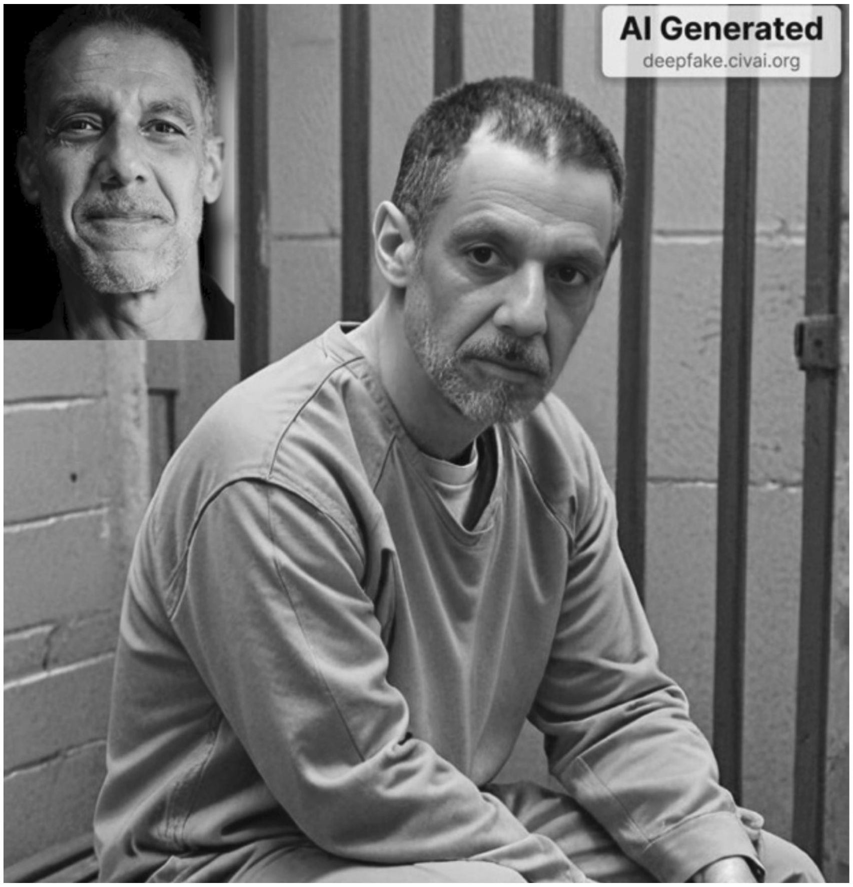


Figure 4: A deepfake in which I inserted my face (source in upper left) into an image of an inmate in an orange jumpsuit.

Starting in 2023, NCEMC has received a small but steadily increasing number of reports that appears to be AI generated or AI manipulated. Given the escalating volume of CSAM reports over the past two decades it was, sadly, predictable that this nascent technology would quickly be weaponized in this horrific way.

18 U.S. Code § 2252A uses a standard prohibiting any visual depiction of CSAM that is "virtually indistinguishable" from a minor engaging in sexual conduct. That is, creation or possession of CSAM can extend beyond material depicting an actual child to material that is computer generated. Beyond the legal standing of AI-generated CSAM, the large-scale creation of abusive content holds the potential to both normalize the sexual abuse of children for offenders and overwhelm an already strained CyberTipline.

While some generative-AI systems placed reasonable guardrails to prevent the creation of CSAM, others did not. Stability AI's first version of their image generator—Stable Diffusion—was open-sourced with no guardrails. In response to concerns of potential abuse, the company's founder, Emad Mostaque, said "Ultimately, it's peoples' responsibility as to whether they are ethical, moral and legal in how they operate this technology." Depending on your viewpoint, this is spectacularly naive, cynical, or simply indifferent.

Fraud

First it was Instagram ads of Tom Hanks promoting dental plans. Then it was TV personality Gayle King hawking a sketchy weight-loss plan. Next, Elon Musk was shilling for the latest crypto scam, and Taylor Swift was announcing a giveaway of Le Creuset cookware. More recently it has been Brad Pitt and Cristiano Ronaldo promoting phony medicines to treat serious diseases like cancer. All, of course, were deepfake scams.

AI-powered scams are not just impacting individuals, they are also impacting small- to large-scale organizations. Earlier this year, a finance worker in Hong Kong was tricked into paying out $25 million to fraudsters using deepfake technology to pose as the company's chief financial officer in a video conference call.

This was not the first such example. In 2019, a United Kingdom based company suffered the same fate when an imposter used an AI-synthesized voice to steal $243,000 in a similar type of scam. And, in early 2020, a United Arab Emirates' bank was swindled out of $35 million when the bank teller was convinced to transfer the funds after receiving a phone call from the purported director of a company whom the bank manager knew and with whom he had previously done business. It was later revealed that the voice was that of an AI-synthesized voice made to sound like the director. These incidents are almost certainly the canaries in the coal mine.

Similar types of fraud are also being carried out at the individual level. In early 2023, the mother of a teenager received a phone call from what sounded like her distressed daughter claiming that the teenager had been kidnapped and feared for her life. The scammer demanded $50,000 to spare the child's life. After calling her husband in a panic, she learned that the daughter was safe at home.

Generative AI is a powerful new weapon in the arsenal of cyber criminals. As synthesized audio and video continue to improve in quality and accessibility, it is reasonable to predict that these technologies will continue to be used to commit a range of small- to large-scale frauds.

Disinformation

By mid-May of 2020, in the midst of the global pandemic, 28 percent of Americans believed Bill Gates planned to use COVID–19 to implement a mandatory vaccine program with tracking microchips. Belief in this conspiracy is not unique to Americans. In global surveys across Central and South America, the Middle East, Northern Africa, the United States, and Western Europe, 20 percent of the public believes this bizarre claim.

As of this year, 22 percent of Americans do not believe in climate change with only 54 percent believing that climate change is human-driven. Understanding of climate change is highly partisan with 93 percent of Democrats and only 62 percent of Republicans believing in climate change.

The far-reaching, far-right QAnon conspiracy alleges a cabal of Satan-worshipping cannibalistic pedophiles is running a global child sex-trafficking ring that was plotting against Donald Trump. A recent poll finds 37 percent of Americans are unsure whether QAnon is true or false, and 17 percent believe it to be true.

Our global health, our planet's health, and our democratic institutions are all under attack due to rampant disinformation, conspiracies, and lies. It seems likely that deepfakes will be an accelerant to disinformation campaigns that until today managed to take significant hold without accompanying visual "evidence."

Liar's Dividend

While the harms from deepfakes are real and already with us, perhaps the most pernicious result of deepfakes and general digital trickery is that when we enter a world where anything we see or hear can be fake, then nothing has to be real. In the era of deepfakes, a liar is equipped with a double-fisted weapon of both spreading lies and using the specter of deepfakes to cast doubt on the veracity of any inconvenient truths—the so-called liar's dividend.

In 2016, for example, Musk was recorded saying "a Model S and Model X at this point can drive autonomously with greater safety than a person. Right now." After a young man died when his self-driving Tesla crashed, his family sued claiming that Musk holds some responsibility because of his claims of safety. In attempting to counter this claim, Musk's attorneys told the court that Musk "like many public figures, is the subject of many 'deepfake' videos and audio recordings that purport to show him saying and doing things he never actually said or did." Fortunately, the judge was not persuaded, "Their position is that because Mr. Musk is famous and might be more of a target for deep fakes, his public statements are immune," wrote Judge Evette Pennypacker. She added, "In other words, Mr. Musk, and others in his position, can simply say whatever they like in the public domain, then hide behind the potential for their recorded statements being a deep fake to avoid taking ownership of what they did actually say and do. The Court is unwilling to set such a precedent by condoning Tesla's approach here."

As deepfakes continue to improve in realism and sophistication it will become increasingly easier to wield the liar's dividend.

Deepfakes: Mitigations

Generative AI continues its ballistic trajectory in terms of its ability to create content that is—or soon will be—nearly indistinguishable from reality. While there are many exciting and creative applications, this technology is also being weaponized against individuals, societies, and democracies.

If we have learned anything from the past two decades of the technology revolution and the disastrous outcomes in terms of invasions of privacy and toxic social media, it is that things will not end well if we ignore, or downplay as the cost of innovation, the malicious uses of generative AI.

I contend that reasonable and proportional interventions from creation through distribution, and across academia, government, and the private sector are both necessary and in the long-term interests of everyone. I will enumerate a range of interventions that are both practical and when deployed properly can keep us safe and allow for innovation to flourish.

Academe

In criticizing the reckless use of scientific advancements without considering the ethical implications, Jeff Goldblum's character, Dr. Ian Malcolm in the 1993 blockbuster movie *Jurassic Park,* said: "Your scientists were so preoccupied with whether they could, they didn't stop to think if they should."

I am, of course, not equating advances in AI with the fictional resurrection of dinosaurs some 66 million years after extinction. The spirit of Goldblum's sentiment, however, is one all scientists should absorb.

Many of today's generative-AI systems used to create NCII and CSAM are derived from academic research. For example, *pix2pix* developed by University of California, Berkeley researchers uses a GAN to transform the appearance or features of an image (*e.g.,* transforming a day-time scene into a night-time scene). Shortly after its release, this open-source software was used to create *DeepNude,* a software that transforms an image of a clothed woman into an image of her unclothed. The creators of *pix2pix* could and should have foreseen this weaponization of their technology and developed and deployed their software with more care.

This was not the first such abuse nor will it be the last. From inception to creation and deployment, researchers need to give more thought on how to develop technologies safely and, in some cases, if the technology should be created in the first place.

1. During the peer-review process, reviewers should assess if any ethical or safety concerns should be considered and/or addressed by the authors prior to publication.
2. While open-source deployments are of great benefit to the larger research community, this benefit should be counter-balanced by the potential risks (as we saw in the above example).
3. At both the undergraduate and graduate levels, mandatory curricular additions are needed to expose math and engineering students to more ethics, history, philosophy, political science, and a broader swath of the liberal arts than most typically see. Our future innovators need the proper scaffolding to think about the broader issues of how technology is intersecting with society and the world beyond Silicon Valley.

Creation

When text-to-image image generators first splashed onto the scene, Google initially declined to release its technology while OpenAI took a more open, and yet still cautious, approach, initially releasing its technology to only a few thousand users. They also placed guardrails on allowable text prompts, including no nudity, hate, violence or identifiable persons. Over time, OpenAI has expanded access, lowered some guardrails and added more features. Stability AI took yet a different approach, opting for a full release of their Stable Diffusion with no guardrails. And most recently Elon Musk's image generator, Grok, followed a similar course leading to all sorts of ridiculous content from Kamala Harris romantically embracing Donald Trump to Mickey Mouse wielding an AR–15, to the more offensive and dangerous.

Regardless of what you think of Google's or OpenAI's approach, Stability AI and Grok made their decisions largely irrelevant: when it comes to this type of shared technology, society is at the mercy of the lowest common denominator. Nevertheless, generative-AI systems should follow several simple rules to mitigate the harm that comes from their services, and the remaining bad actors will have to be dealt with through legislation and litigation (see below).

1. The Coalition for Content Provenance and Authentication (*https://c2pa.org*) is a multi-stake holder, open-source initiative aimed at establishing trust in digital audio, image, and video. The focus of the C2PA is creating standards to ensure the authenticity and provenance of digital content. This standard includes the addition of metadata and embedding an imperceptible watermark into content, and extracting a distinct digital signature from content that can identify content even if the attached content credentials are stripped out. Any AI-generated service should implement this standard to make it easier to identify content as AI-generated.
2. Because text-to-image and text-to-video systems are capable of producing content limited only by the imagination of the creator, some reasonable semantic guardrails should be implemented on both the input and output. On the input side, a large language model (LLM) can flag prompts that includes requests for NCII, CSAM, or other violative or illegal content. On the output side a multimodal LLM can similarly flag violative content that managed to slip through the input guardrails.
3. Although content credentials and semantic guardrails are important steps in mitigating harms, they are not infallible. Generative-AI services should adopt a know your customer (KYC) approach common in all financial institutions. This will both put creators on notice that their content creation is not anonymous, and allow platforms to aid investigations into illegal uses of their services.

Distribution

There are three main phases in the life cycle of online content: creation, distribution, and consumption. I have addressed creation in the previous section and will address consumption next. On the distribution side, social media needs to take more responsibility for everything from the unlawful to the lawful-but-awful content that is both shared on their platforms and amplified by their own recommendation algorithms.

While it is easy to single out social media platforms for their failure to rein in the worst abuses on their platforms from CSAM, to NCII, fraud, violence, and dangerous disinformation campaigns, these platforms are not uniquely culpable. Social media operates within a larger online ecosystem powered by advertisers, financial services, and hosting/network services.

Each of these—often hidden—institutions must also take responsibility for how their services are enabling a plethora of online harms.

1. In addition to improving on their content moderation policies and enforcement, social media can create a global shared database of identified NCII as they have previously done for CSAM and terror-related content. Once NCII is identified, such a shared database would prevent NCII from being re-uploaded thus reducing the continued harm to victims.
2. Pressure to effect change on platforms rarely comes from users because we are not the customer, we are the product. The real customers are advertisers who should wield their power to effect change by insisting, for example, that their products and services not be advertised along side CSAM, NCII, and violent content. This isn't just the right thing to do, it is the smart thing to do for brand protection.
3. The largest financial services (Visa, MasterCard, PayPal, etc.) should not be in business with services that primarily host or produce NCII or other illegal and harmful content. There are at least two examples of where financial services were able to effect change when they withheld service from PornHub (for hosting CSAM and NCII) and Backpage (for enabling sex trafficking).
4. While more fraught, computing infrastructure services from GitHub to Amazon/Google/Microsoft cloud, to network services like Cloudflare can also act as better stewards. For example, at a hate-filled neo-Nazi march in Charlottesville, Virginia in 2017, violence erupted between marchers and counter protesters leading to the horrific murder of a counter-protester. In the aftermath, companies like Cloudflare came under heavy criticism for providing services to neo-Nazi groups like Daily Stormer, and for giving them personal information on people who complain about their content. Despite initially refusing to act, Cloudflare eventually terminated the account of Daily Stormer. While these groups will eventually find another home, that doesn't mean that we should not continually make the inter-net—where they can amplify their hate and violence—an increasingly unwelcome place.

Consumption

When discussing deepfakes, the most common question I'm asked is "how can the average consumer distinguish the real from the fake?" My answer is always the same: "nothing." After which I explain that artifacts in today's deepfakes—seven fingers, incoherent text, mismatched earrings, etc.—will be gone tomorrow, and my instructions will have provided the consumer with a false sense of security. The space of generative AI is moving too fast and the forensic examination of an image is too complex (see next section) to empower the average consumer to be an armchair forensic detective.

There are, however, things that consumers can do to protect themselves from the being defrauded or fooled by deepfakes.

1. Protecting against fraudulent phone calls from scammers claiming to be a family member can be as simple as having an agreed upon family code word that would have to be produced when an unexpected or emergency call is received.
2. Protecting against disinformation is, of course, more challenging as more and more people get the majority of their news from increasingly louder echo chambers. Here, I propose the development of a national K–12 effort to educate students on how to strike a balance between skepticism and vigilance, how to spot signs of disinformation, how to fact check, and how to generally be better digital citizens than the previous generation.
3. Protecting against being a victim of NCII effectively requires being completely invisible online, which in today's world is nearly impossible. If you are a victim of NCII, several organizations may be able to provide assistance or advice, including the Cyber Civil Rights Initiative (*https://cybercivilrights.org*).

Legislation

Existing legislation should be sufficient to combat child sexual abuse material (CSAM) and fraud, weather AI-powered or not. Here interventions to protect the public and prosecute perpetrators are primarily limited by law enforcement resources and the inaction of the largest social media platforms. With tens of millions of CSAM reports each year to NCMEC, for example, law enforcement is simply overwhelmed. With billions of uploads each day to social media, these platforms are incapable—and too often unwilling—to combat illegal activity on their services.

Most agree that bans or restrictions should be placed on the creation and distribution of non-consensual intimate imagery (NCII), but the law has not fully caught up with the latest technology that now makes it too easy to create and distribute this type of content.

In recent years, however, there has been a patchwork of national and international legislation enacted. In 2019, the U.S. state of Virginia expanded its 2014 "revenge porn" laws to include synthesized or manipulated content, making it illegal to share nude photos or videos of anyone—real or fake—without their permission. California, Hawaii, New York, and Texas have similar restrictions, but as of yet there is no Federal legislation. In 2021, Australia amended its laws to include synthesized or manipulated content; violations can incur both criminal charges and monetary fines.

Because the Internet is borderless, nations should now band together to move from a patchwork of legislation to a consistent set of rules and regulations to combat NCII. It remains unclear, however, whether legislation can fully rein in these abuses. Hollywood actress Scarlett Johansson—a frequent target of NCII—told the *Washington Post,* "I think it's a useless pursuit, legally, mostly because the Internet is a vast wormhole of darkness that eats itself."

With some exceptions including speech designed to interfere with elections or the peaceful transfer of power, mitigating the harms from various forms of political speech is complex, and legally fraught. It is not, after all, illegal for a politician to lie or for anyone to believe those lies.

Nevertheless, several states have recently passed legislation designed to protect the integrity of elections from misleading deepfakes. In 2024, in the lead up to a contentious national election in which deepfakes have already played a roll, both the states of Minnesota and California passed legislation to impose varying civil and criminal penalties to those creating, distributing, and in some cases, hosting, AI-powered election misinformation. These laws are not without controversy and they will soon be challenged on First Amendment grounds.

Nevertheless, practical and proportional responses to existing and emerging threats are within reach.

1. Despite Scarlett Johansson's perfectly reasonable assessment of the state of the internet, a combination of updating of existing legislation and crafting new leg-

islation to combat emerging threats is necessary, if not sufficient. To date, only a handful of nations and a handful of U.S. states have moved to mitigate the harms from deepfakes. While I applaud individual U.S. states for their efforts, Internet regulation cannot be effective with a patchwork of state laws. A national and coordinated international effort is required. In this regard, the European Union's *Digital Safety Act,* the United Kingdom's *Online Safety Act,* and Australia's *Online Safety Act* provide a road map for the U.S. While regulation at a global scale will not be easy, some common ground can surely be found among the U.S. and it allies, thus serving as a template for other nations to customize and adopt.

2. In the absence of sweeping legislation, liability can be a powerful motivating factor for the technology sector to make sure their products and services are not harmful. But, penetrating the powerful liability shield of Section 230 of the Communications Decency Act, has proven challenging. Written in 1996, Section 230 provides broad immunity to online platforms (including social media) from being held liable for user-generated content, and it allows platforms to moderate content in "good faith" without being treated as the publisher or speaker of the content. The U.S. Congress has repeatedly tried (and failed) to modernize this outdated law that could not of and does not work in today's modern technology landscape. The U.S. Congress needs to revisit the issue by modernizing Section 230 to create some liability to motivate a mindset of safety by design, not safety as (at best) an afterthought.
3. On the specific issues of CSAM and NCII, more resources and training should be provided to law enforcement to provide resources for victims and support for investigations and, where appropriate, prosecutions.

Summary

There is much to be excited about in this latest wave of the technology revolution. But, if the past few technology waves have taught us anything, it is that left unchecked, technology will begin to work against us and not for or with us. We need not make the mistakes of the past. We are nearing a fork in the road for the type of future we want and what role technology will play.

Famed actor and filmmaker Jordan Peele's 2018 public service announcement on the dangers of fake news and the then-nascent field of deepfakes[3] offers words of advice and caution. The PSA concludes with a Peele-controlled President Obama saying "how we move forward in the age of information is gonna be the difference between whether we survive or whether we become some kind of f****d up dystopia." I couldn't agree more.

Senator HICKENLOOPER. Thank you, Dr. Farid.

Mr. Brookman, you are next, the technology—Director of Technology Policy for Consumer Reports—and the first subscription I ever made to a magazine was *Consumer Reports* just because I thought that the consumers needed someone to stick up for them, and someone told me recently that there are now 93,000 trade associations. As far as I know, I think there is only one *Consumer Reports.* So——

STATEMENT OF JUSTIN BROOKMAN, DIRECTOR, TECHNOLOGY POLICY, CONSUMER REPORTS

Mr. BROOKMAN. Thank you very much for that intro, Senator, and thank you very much for the opportunity to testify here today.

As you say, I am here from Consumer Reports where I head up our work on tech policy advocacy. We are the world's largest independent testing organization. We use our ratings, our journalism, our survey, and our advocacy to try to argue for a better, fairer marketplace for consumers.

This hearing is about AI-enabled fraud but I do want to say a couple words about the very real consumer benefits from AI. Cer-

[3] Deepfake Obama: https://www.youtube.com/watch?v=cQ54GDm1eL0

tainly, there have been a number of tremendous scientific advances.

We are seeing every day tangible real-world benefits to it—autonomous taxis in San Francisco, real-time language translation. These are amazing advances that do make consumers' lives better.

However, the widespread availability of AI tools does lead to some very real harms. Companies can now collect, process, share more information about us, giving them more power over us including the power to do personalized pricing based on how much we are willing to pay.

It can rip off content creators who see the products of their labor turned into AI slop of dubious value and it can let people create nonconsensual intimate images of acquaintances, celebrities, or classmates.

Another harm is that it makes—AI makes it a lot easier for scammers to rip people off and take their money. It makes targeted spear phishing attacks a lot more efficient.

Spammers can spin up hundreds, thousands, of personalized solicitations using generative AI tools. This is the scale that Senator Blackburn was talking about.

It lets companies generate dozens of fake reviews with just a few keystrokes. Review fraud is already a massive, terrible problem online. AI just makes it easier to scale and flood the zone.

One of the most dangerous tools is the one that you talked about, Senator Hickenlooper—realistic voice impersonation. Scammers get a clip of someone's voice, take it to a company like ElevenLabs, and then get it to follow a script.

There have been numerous examples of people calling family members with some degree of urgency saying, "I have been in an accident" or calling a co-worker and said, "I need you to wire money to us."

We have heard from dozens of Consumer Reports members who say they got these calls and it sounded—you know, from family members, sounded incredibly convincing. Some actually lost money and even though they felt like they were savvy, smart people.

These tools are really easy to find online. You can search Google for voice impersonation tools to get lots of options including sponsored results who pay Google to get higher up in those results.

Consumer Reports is in the middle of evaluating some of these companies and a lot just do not have any or very few protections in place to stop fraudulent uses like requiring someone to read a script so you know this person is consenting to have their voice used. A lot of them do not have that.

So what should be done? In my written testimony I lay out five areas where I think we need to see some improvements. One, stronger enforcers. The Federal Trade Commission did the impersonation rule that was talked about. They did bring five important cases. That is part of Operation AI Comply in September.

Those are great but they just do not have the capacity today to deal with this wave of AI-enabled fraud. The FTC is understaffed. They cannot give statutory penalties for the last several years. They often cannot even get scammers to give up the money they have stolen. These are problems that need to be addressed.

Tool and platform responsibilities—some of these AI-powered tools are designed such that they are mostly going to be used for illegitimate purposes, whether it is through the creation of deepfake intimate images or voice impersonation.

These companies should have heightened obligations to try to forestall harmful uses. If they cannot do that maybe they should not be publicly or so easily available.

General purpose AI can be harmful, too, but it is going to be harder for those companies to think about every possible negative use case and more expensive and may forestall good uses.

But it does seem like in many cases they should be doing more. You can go to ChatGPT today and say, hey, create 20 fake reviews for my restaurant and ChatGPT will say, sure, here are 20 fake reviews and here are my favorite dishes and here is what great service I had, which, again, can help companies flood the zone with fake reviews.

Transparency—people deserve to know when they are interacting with a real person or when they are seeing fake content. I know there has been a lot of bills introduced in this Congress including—and legislation passed in California—A.B. 982—that would require some advances there.

Fourth, stronger privacy and security laws. The United States in general has very weak protections, as Senator Blackburn talked about. We have seen a lot of progress at the state level in recent years. These laws still are not strong enough.

And then five, user education and better tools. I do not want to put all the burden on consumers but this is the reality of the world we live in. We need to educate people about what to expect.

We are part of a campaign called "Take Nine" that tries to teach people when you get some sort of urgent call to action, pause, take 9 seconds, think about it—that is real or a scam.

And over time the tools need to improve, too. I look forward to hearing from some of the other witnesses about tools that are advancing and, hopefully, these can be built into the platforms that we use to access the internet.

Thank you very much for your time and I look forward to answering your questions.

[The prepared statement by Mr. Brookman follows:]

Prepared Statement of Justin Brookman, Director, Technology Policy, Consumer Reports

On behalf of Consumer Reports, I want to sincerely thank you for the opportunity to testify here today. We appreciate the leadership of Chairman Hickenlooper and Ranking Member Blackburn not only for holding this important hearing, but also for working in a constructive, bipartisan fashion to develop smart and effective policy solutions to protect American consumers from increasingly sophisticated fraud and scams powered by artificial intelligence.

Founded in 1936, Consumer Reports (CR) is an independent, nonprofit, and nonpartisan organization that works with consumers to create a fair and just marketplace. Known for its rigorous testing and ratings of products, CR advocates for laws and company practices that put consumers first. CR is dedicated to amplifying the voices of consumers to promote safety, digital rights, financial fairness, and sustainability. The organization surveys millions of Americans every year, reports extensively on the challenges and opportunities for today's consumers, and provides ad-free content and tools to our six million members across the United States.

I have been head of Technology Policy for Consumer Reports for seven years, and during all that time we have been active on AI policy.[1] We have called for stronger privacy protections for consumers' data even before the widespread advent of AI, back when the buzzword was "Big Data" instead.[2] We have supported federal[3] and state[4] legislation and rulemaking[5] to require developers of automated decision-making systems to provide consumers information about how those systems work and to account for potential bias.[6] We have written about the importance of independent testing of AI systems, calling on policymakers to make changes to existing laws that often impede good faith research.[7] And of course, Consumer Reports has long been active on scams as well, offering tools to educate consumers on how to protect themselves,[8] and petitioning Congress to give the Federal Trade Commission the tools it needs to more aggressively pursue wrongdoers.[9]

In my testimony today, I will discuss the benefits and risks to consumers from the widespread use of artificial intelligence, including detailing how AI is fuelling increasingly sophisticated scams that cost consumers billions of dollars a year. I will then discuss existing legal protections and consumer education efforts that have advanced significantly in recent years but which have still proved insufficient to the scope of the problem. Finally, I will discuss potential solutions including:

- Stronger enforcement bodies,
- Clearer platform accountability rules,
- Transparency obligations,
- Stronger privacy and security laws, and

[1] Katie McInnis, *Pre-Hearing Comments on Consumer Privacy for the Federal Trade Commission's Hearings on Competition and Consumer Protection in the 21st Century on February 12–13, 2019, FTC–2018–0098,* Consumer Reports Advocacy, (Dec. 21, 2018), *https://advocacy.consumerreports.org/wp-content/uploads/2018/12/Consumer-Reports-comments-FTC-2018-0098-2.pdf.*

[2] Press Release, *Consumer Reports Launches Digital Standard to Safeguard Consumers' Security and Privacy in Complex Marketplace,* Consumer Reports, (Mar. 6, 2017), *https://www.consumerreports.org/media-room/press-releases/2017/03/consumer_reports_launches_digital_standard_to_safeguard_consumers_security_and_privacy_in_complex_marketplace/.*

[3] Press Release, *Senator Markey Introduces AI Civil Rights Act to Eliminate AI Bias, Enact Guardrails on Use of Algorithms in Decisions Impacting People's Rights, Civil Liberties, Livelihoods,* Ed Markey United States Senator for Massachusetts, (Sep. 24, 2024), *https://www.markey.senate.gov/news/press-releases/senator-markey-introduces-ai-civil-rights-act-to-eliminate-ai-bias-enact-guardrails-on-use-of-algorithms-in-decisions-impacting-peoples-rights-civil-liberties-live lihoods.*

[4] Grace Gedye, *Consumer Reports backs signing of high-risk AI bill, calls on Colorado General Assembly to strengthen it before it goes into effect,* Consumer Reports, (May 18, 2024), *https://advocacy.consumerreports.org/press_release/consumer-reports-backs-signing-of-high-risk-ai-bill-calls-on-colorado-general-assembly-to-strengthen-it-before-it-goes-into-effect/.* Gedye currently serves on the Colorado Artificial Intelligence Impact Task Force set up to make recommendations to the Colorado legislature to revise the law before it goes into effect in 2026. *See* Artificial Intelligence Impact Task Force, Colorado General Assembly, *https://leg.colorado.gov/committees/artificial-intelligence-impact-task-force/2024-regular-session.*

[5] Matt Schwartz and Justin Brookman, *Consumer Reports Submits Comments on the California Privacy Protection Agency's Preliminary Rulemaking on Cybersecurity Audits, Risk Assessments, and Automated Decisionmaking, Consumer Reports Advocacy,* (Mar. 27, 2023), *https://advocacy.consumerreports.org/research/consumer-reports-submits-comments-on-the-california-privacy-protection-agencys-preliminary-rulemaking-on-cybersecurity-audits-risk-assessments-and-automate d-decisionmaking;* Justin Brookman *et al., Consumer Reports submits comments on FTC privacy and security rulemaking,* Consumer Reports Advocacy, (Nov. 21, 2022), *https://advocacy.consumerreports.org/research/consumer-reports-submits-comments-on-ftc-privacy-and-security-rulemaking/.*

[6] Grace Gedye and Matt Scherer, *Opinion* | Are These States About to Make a Big Mistake on AI?, Politico, (Apr. 30, 2024), *https://www.politico.com/news/magazine/2024/04/30/ai-legislation-states-mistake-00155006.*

[7] Nandita Sampath, *New Paper: Opening Black Boxes: Addressing Legal Barriers to Public Interest Algorithmic Auditing,* Consumer Reports Innovation Blog, (Oct. 13, 2022), *https://innovation.consumerreports.org/new-paper-opening-black-boxes-addressing-legal-barriers-to-public-interest-algorithmic-auditing/.*

[8] Security Planner, Consumer Reports, *https://securityplanner.consumerreports.org/.*

[9] Letter from Consumer Reports to Chairwoman Rosa L. DeLauro *et al.,* (May 25, 2021), *https://advocacy.consumerreports.org/wp-content/uploads/2021/05/CR-letter-on-FTC-appropriations-052521.pdf* (petitioning for an increase in funding for the FTC); Testimony of Anna Laitin, Director, Financial Fairness and Legislative Strategy, Consumer Reports Before the House of Representatives Committee on Energy & Commerce Subcommittee on Consumer Protection and Commerce on "The Consumer Protection and Recovery Act: Returning Money to Defrauded Consumers," (Apr. 27, 2021), *https://www.congress.gov/117/ meeting/house/112501/witnesses/HHRG-117-IF17-Wstate-LaitinA-20210427.pdf* (petitioning for the restoration of the FTC's 13(b) injunctive authority).

- Citizen education and better tools

I. The Very Substantial Benefits and Risks of AI

As an initial matter, we must recognize the massive societal benefits from the advent of artificial intelligence, which can accomplish a broad variety of important tasks far more efficiently than traditional methods. Earlier this month, an MIT professor published a paper detailing the benefits to materials research from the use of AI, leading to a 44 percent increase in materials, a 39 percent increase in patent filings, and a 17 percent rise in downstream innovation.[10] Even when it comes to scams, AI does and will continue to play an important defensive role, improving spam filters and search engine ranking, identifying bad actors, and alerting consumers to potentially fraudulent solicitations.[11]

We use artificial intelligence at Consumer Reports in a variety of ways to make us more effective in our mission to deliver a fairer, safer marketplace for consumers. In our testing on privacy and security, we use AI to automate document collection and policy review to speed up product evaluations. We have used machine learning to analyze large data sets to find evidence of racial discrimination in auto insurance prices.[12] We are looking to use AI semantic tools to expand our early warning system to monitor publicly available product reviews to detect potentially dangerous or defective products. And earlier this year we rolled out "AskCR"—a generative-AI system designed to more effectively draw upon CR's extensive data troves to provide answers to our members' questions about various products.[13]

However, artificial intelligence, like any tool, can be used for harm as well. And artificial intelligence is a very powerful tool. It can scrape and steal content from publicly available sources, depriving content creators of the value of their work and substituting it with AI-generated slop of dubious provenance.[14] AI can exacerbate privacy invasions, giving companies more data and power over us and the ability to personalize prices to extract greater proportions of consumer surplus from any transaction.[15] AI makes it easy to generate nonconsensual sexual images just by uploading a picture of an acquaintance or celebrity.[16] AI could also lead to increased corporate consolidation and perversely less innovation if only the companies with the most existing resources can take advantage of technological advances.[17] These are very real harms not solved by the existing marketplace and they need serious policy solutions.

The use of AI to power fraud and scams

Artificial intelligence is also a useful tool for fraud and scams, automating previously laborious tasks and sometimes enabling new capabilities entirely. Recent research suggests that generative AI can be used to scale "spear phishing"—the personalization of phishing messages based on personal data to make them more convincing. By using freely available generative AI services, researchers found the cost of creating individualized spear phishing solicitations fell from $4.60 to just 12 cents

[10] Aidan Toner-Rodgers, *Artificial Intelligence, Scientific Discovery, and Product Innovation,* GitHub, (Nov. 6, 2024), *https://aidantr.github.io/files/AI_innovation.pdf.*

[11] Fredrik Heiding *et al., Devising and Detecting Phishing E-mails Using Large Language Models,* IEEE Explore, (Mar. 11, 2024), *https://ieeexplore.ieee.org/document/10466545.*

[12] Jeff Larson *et al., How We Examined Racial Discrimination in Auto Insurance Prices,* ProPublica and Consumer Reports, (Apr. 5, 2017), *https://www.propublica.org/article/minority-neighborhoods-higher-car-insurance-premiums-methodology.*

[13] AskCR, Consumer Reports, *https://innovation.consumerreports.org/initiatives/askcr/.*

[14] Benjamin Hoffman, *First Came 'Spam.' Now, With A.I., We've Got 'Slop',* New York Times, (Jun. 11, 2024), *https://www.nytimes.com/2024/06/11/style/ai-search-slop.html.*

[15] Brian Pearson, Personalizing Price With AI: How Walmart, Kroger Do It, Forbes, (Sep. 7, 2021), *https://www.forbes.com/sites/bryanpearson/2021/09/07/personalizing-price-with-ai-how-walmart-kroger-d o-it/.* Another way AI can lead to higher consumer prices is when multiple sellers use the same algorithm to help set prices. Using the nonpublic data from all its customers together, the AI vendor can recommend to all its customers universally higher prices, especially in markets where a greater number of market participants use its systems. *See* Hannah Garden-Monheit and Ken Merber, Price fixing by algorithm is still price fixing, Federal Trade Commission Business Blog, (Mar. 1, 2024), *https://www.ftc.gov/business-guidance/blog/2024/03/price-fixing-algorithm-still-price-fixing.* Indeed, some academics have suggested that even different algorithms based on different data may implicitly collude if both are independently setting prices, leading to higher costs from consumers. *See* Ariel Ezrachi and Maurice Strucke, Sustainable and Unchallenged Algorithmic Tacit Collusion, 17 Northwestern Journal of Technology and Intellectual Property 217 (2020).

[16] Matteo Wong, *High School Is Becoming a Cesspool of Sexually Explicit Deepfakes,* The Atlantic, (Sep. 26, 2024), *https://www.theatlantic.com/technology/archive/2024/09/ai-generated-csam-crisis/680034/.*

[17] Jai Vipra and Anton Korinek, *Market concentration implications of foundation models: The Invisible Hand of ChatGPT,* Brookings, (Sep. 7, 2023), *https://www.brookings.edu/articles/market-concentration-implications-of-foundation-models-the-invisible-hand-of-chatgpt/.*

per message.[18] AI allows fraudsters to spin up fake websites with just a few clicks that look like legitimate services.[19] AI could also help bad actors supercharge search engine optimization efforts to fool search engines into displaying fake customer service numbers for popular companies. (The *Washington Post* recently reported that scammers were easily able to get Google to provide fake phone numbers for companies such as Delta and Coinbase).[20]

One area where Consumer Reports has focused its research is on the use of AI for voice cloning. AI voice cloning tools have the potential to supercharge impersonation scams, including a phone scam sometimes known as the 'Grandparent scam', in which a consumer is contacted and is told that a loved one is in trouble: they wrecked their car or they landed in jail and need money fast.[21] In the past, scammers might try to achieve a rough approximation of a young relative's voice. Now, if scammers have access to audio of a family member speaking, from, say, social media videos, they can create a potentially compelling AI clone of their voice.

This is already happening. The *Washington Post* has covered consumers who sent thousands of dollars to scammers after thinking they've heard their family member on the phone in need of help.[22] The *New Yorker* highlighted the stories of several parents sent into a state of terror after thinking they heard their panicked child's voice, followed by dark threats.[23] Scammers have targeted companies as well, using AI voice tools to convince employees they are getting a call from a higher up who needs them to transfer funds. In one case, the managing director of a British energy company wired $240,000 to Hungary, thinking he was speaking to his boss.[24]

CR reached out to consumers across the country in February, asking if they had received a phone call from a scammer mimicking the voice of someone they knew, or someone well-known. We heard from consumers who said the experience left them feeling "vulnerable," "shaken by the experience," and "really weirded out":[25]

- "My Grandpa got a call from someone claiming to be me. Supposably, I was traveling, and my car broke down and I needed to have him send money so I could complete my travels. Grandpa said there was no doubt in his mind that I was the caller and was preparing to do as asked. Luckily, before he went through with the transaction, he reasoned that if I was in trouble and honestly needed money, he would have heard from my mom. . . . Scary that the tools they use could imitate my voice that closely as to fool a close relative. . ."—member from Minnesota
- "The initial caller's voice sounded very much like my nephew's. He knew family details, pleaded with me not to call his father and promised to pay me back as soon as he got home—all very convincing. I should add that I spent more than 60 years in law-enforcement and intelligence work. This scam was so carefully arranged and executed that I fell for it nevertheless."—member from Massachusetts
- "I received a phone call from my grandson explaining that he was in a car accident at college and needed $5,000. He sounded scared and upset and asked that I not tell his parents. So, I went to my bank to get the money and the bank teller told me it was a scam. I did not believe her as I was sure it was my grandson's voice."—member New York

[18] Fredrik Heiding *et al., Devising and Detecting Phishing E-mails Using Large Language Models,* IEEE Explore, (Mar. 11, 2024), *https://ieeexplore.ieee.org/document/10466545.*

[19] Chris Smith, *Mind-blowing AI instantly makes artificial web pages from anything you type,* Boy Genius Report, (Jun. 26, 2024), *https://bgr.com/tech/mind-blowing-ai-instantly-makes-artificial-web-pages-from-anything-you-type/.*

[20] Shira Ovide, *Don't trust Google for customer service numbers. It might be a scam.,* Washington Post, (Aug. 20, 2024, *https://www.washingtonpost.com/technology/2024/08/20/google-search-scams-customer-service-phone-numbers/.*

[21] Consumer Alert, *Scammers use AI to enhance their family emergency schemes,* Federal Trade Commission, (Mar. 20, 2023), *https://consumer.ftc.gov/consumer-alerts/2023/03/scammers-use-ai-enhance-their-family-emergency-sch emes.*

[22] Pranshu Verma, *They thought loved ones were calling for help. It was an AI scam.,* Washington Post, (Mar. 5, 2023), *https://www.washingtonpost.com/technology/2023/03/05/ai-voice-scam/.*

[23] Charles Bethea, *The Terrifying A.I. Scam That Uses Your Loved One's Voice,* New Yorker, (Mar. 7, 2024), *https://www.newyorker.com/science/annals-of-artificial-intelligence/the-terrifying-ai-scam-that-uses-your-loved-ones-voice.*

[24] Drew Harwell, *An artificial-intelligence first: Voice-mimicking software reportedly used in a major theft,* Washington Post, (Sep. 4, 2019), *https://www.washingtonpost.com/technology/2019/09/04/an-artificial-intelligence-first-voice-mimicking-software-reportedly-used-major-theft/*
.

[25] Member Stories, *Have you received a phone call that impersonated someone?,* Consumer Reports, *https://www.consumerreports.org/stories?questionnaireId=307.*

- “The voice on the other end sounded just like my grandson and it said ‘Gramie, I’ve been in an accident.’ ”—member from Florida
- “I was skeptical, and told him I had heard of scams such as this. So he said, ‘I’ll let Nate say a few words to you.’ It sounded exactly like my Nate!! He has a rather unusual voice, so I was then almost convinced.”—member in Indiana
- “The phone rang and a voice said, ‘Hi Gramma, this is Mac. I’m in New Jersey with my friend Chris. We had an accident. I broke my nose.’ I immediately knew it wasn’t my grandson. He calls me Gramma Beth . . . and he’d have no reason to be in New Jersey. He’s New York, born and bred . . . The voice did sound exactly like him, however, and I could easily have been duped.”—member from New York
- “I received a call and heard my daughter crying hysterically! She wasn’t making sense so an ‘officer’ took over the call. He stated I needed to come right away but would not answer my questions. Thankfully I have Life360 and looked to see where my daughter was at and it showed her at home. . . . To hear my daughter’s crying voice shook me for a long time!”—member from Minnesota

Imposter scams are common. In 2023, 853,935 imposter scams were reported to the FTC’s Consumer Sentinel Network.[26] Twenty one percent of those scam reports included monetary losses, which totalled $2.7 billion in 2023.[27] Imposter scams were the second most frequently reported category of fraud reported to the Consumer Sentinel Network in 2023, out of the 29 categories that the network tracks.[28]

Consumer Reports recently conducted a nationally representative survey of consumers and asked about scams. Nearly nine percent of respondents said they had lost money due to a cyberattack or digital scam.[29] The numbers were even more stark for Black and Hispanic consumers; for those consumers who had encountered a scam attempt, 33 percent of Black consumers and 30 percent of Hispanic consumers had lost money as result.[30]

Increasing use of payment apps such as Zelle and the rise of cryptocurrency also have fuelled scams. These payment methods lack the statutory protections of traditional banking and credit card payments, so consumers who find themselves defrauded may lack any practical recourse to get their money back.[31] Many consumers may not know that these newer payment methods lack legal protections: A 2022 Consumer Reports evaluation of four leading peer-to-peer payment apps showed that all four lacked clear, accessible disclosures about the availability of FDIC insurance for user balances and the full scope of their fraud and error resolution policies.[32]

AI voice cloning tools can also be used for deception with different ends—including harming someone’s reputation. A Maryland high school athletic director reportedly used AI voice cloning tools to mimic the voice of a school principal.[33] The recording came after the athletic director and the principal had discussed the athletic director’s poor work performance. The manufactured audio clip reportedly contained racist remarks about Black students’ test taking abilities, as well as antisemitic comments.

Fake endorsements and reviews

AI voice and likeness cloning tools have unlocked scammers’ abilities to generate deepfake videos falsely depicting celebrities and political figures endorsing products,

[26] Consumer Sentinel Network Data Book 2023, Federal Trade Commission, (Feb. 2024), at 4, 7, *https://www.ftc.gov/system/files/ftc_gov/pdf/CSN-Annual-Data-Book-2023.pdf.*

[27] *Id.* at 4.

[28] *Id.* at 7.

[29] Yael Grauer, *New Report: 2024 Consumer Cyber Readiness,* Consumer Reports Innovation Blog, (Oct. 1, 2024), at 4 *https://innovation.consumerreports.org/new-report-2024-consumer-cyber-readiness/.*

[30] *Id.* at 4; *see also* Report to Congress, Combating Fraud in African American & Latino Communities The FTC’s Comprehensive Strategic Plan, Federal Trade Commission, (Jun. 15, 2016), *https://www.ftc.gov/system/files/documents/reports/combating-fraud-african-american-latino-communities-ftcs-comprehensive-strategic-plan-federal-trade/160615fraudreport.pdf.*

[31] Testimony of Delicia Reynolds Hand, Senior Director, Digital Marketplace, Consumer Reports, Before the United States Senate Permanent Subcommittee on Investigations on “Fraud Alert!: Shedding Light on Zelle,” (May 21, 2024), *https://advocacy.consumerreports.org/ wp-content/uploads/2024/05/Fraud-Alert_-Shedding-Light-on-Zelle.Delicia-Hand.May-2024.pdf.*

[32] Delicia Hand, *Peer-to-Peer Payment Apps: A Case Study for a Digital Finance Standard,* Consumer Reports Advocacy, (Jan. 24, 2023), *https://advocacy.consumerreports.org/research/peer-to-peer-payment-apps-a-case-study-for-a-digital-finance-standard/.*

[33] Ben Finley, *Athletic director used AI to frame principal with racist remarks in fake audio clip, police say,* AP News, (Apr. 25, 2024), *https://apnews.com/article/ai-artificial-intelligence-principal-audio-maryland-baltimore-county-pikesville-853ed171369bcbb888eb54f55195cb9c.*

suggesting investments, and urging citizens to take action. Recent research suggests that consumers struggle to recognize deepfake videos as false, and also overestimate their own ability to detect deepfakes.[34]

AI-powered celeb-bait has proliferated on social media. An investigation by *ProPublica* identified videos on Meta seemingly depicting President-elect Trump and President Biden—each with their distinctive tone and cadence—offering cash handouts if people filled out an online form.[35] *404 Media* has reported on the spread of low-rent AI clones of Joe Rogan, Taylor Swift, Ice Cube, Andrew Tate, Oprah, and The Rock pushing Medicare and Medicaid-related scams on YouTube.[36] Scammers have used an AI clone of Taylor Swift's to hawk Le Creuset dishware.[37] Elon Musk's likeness and voice has been frequently repurposed by scammers using AI video and voice tools to push fraudulent "investment" schemes. One consumer was reportedly scammed out of $690,000 after seeing a deepfaked Elon Musk endorse an investment opportunity.[38]

AI can also make it easier to illegally promote products through the creation of mass fake reviews. Biased and downright fraudulent reviews are rampant online. Popular sites are riddled with thousands of dubious reviews, polluting the information available to consumers to make an informed choice.[39] One study finds that nearly half of reviews for clothes and apparel are faked—and, on average across all product lines, 39 percent of the reviews are false.[40] Another study put the number at closer to 30 percent.[41] And another found that online reviews are, overall, untrustworthy through a variety of metrics, including convergence with Consumer Reports ratings and resale value.[42]

Using generative AI, a fraudster can generate dozens of realistic sounding fake reviews in seconds. Inputting into ChatGPT for example the prompt "generate ten fake five star reviews of varying length and tone for the Ukrainian DC restaurant Ruta" results in ChatGPT's response: "Here are ten fake five-star reviews for the Ukrainian restaurant Ruta, showcasing a variety of tones and lengths:" followed by ten detailed reviews praising particular dishes, the decor, and the service. Earlier this year, the FTC brought a case against the generative AI service Rytr for offering a product that would generate unlimited reviews for a product or service with limited user input—Rytr would then create detailed reviews with invented details and anecdotes.[43]

[34] Nils C Köbis *et al., Fooled twice: People cannot detect deepfakes but think they can,* National Library of Medicine National Center for Biotechnology Information, (2021), *https://pubmed.ncbi.nlm.nih.gov/34820608/.*

[35] Craig Silverman and Priyanjana Bengani, *Exploiting Meta's Weaknesses, Deceptive Political Ads Thrived on Facebook and Instagram in Run-Up to Election,* ProPublica, (Oct. 31, 2024), *https://www.propublica.org/article/facebook-instagram-meta-deceptive-political-ads-election.*

[36] Jason Koelber, *Deepfaked Celebrity Ads Promoting Medicare Scams Run Rampant on YouTube,* 404 Media, (Jan. 9, 2024), *https://www.404media.co/joe-rogan-taylor-swift-andrew-tate-ai-deepfake-youtube-medicare-ads/.*

[37] Tiffany Hsu and Yiwen Lu, *No, That's Not Taylor Swift Peddling Le Creuset Cookware,* New York Times, (Jan. 9, 2024), *https://www.nytimes.com/2024/01/09/technology/taylor-swift-le-creuset-ai-deepfake.html.*

[38] Stuart Thompson, *How 'Deepfake Elon Musk' Became the Internet's Biggest Scammer,* New York Times, (Aug. 14, 2024), *https://www.nytimes.com/interactive/2024/08/14/technology/elon-musk-ai-deepfake-scam.html.*

[39] Simon Hill, *Inside the Market for Fake Amazon Reviews,* Wired, (Nov. 2, 2022), *https://www.wired.com/story/fake-amazon-reviews-underground-market/;* Joe Enoch, *Can You Trust Online Reviews? Here's How to Find the Fakes,* NBC News (Feb. 27, 2019), *www.nbcnews.com/business/consumer/can-you-trust-online-reviews-here-s-how-find-fakes-n976756.*

[40] Eric Griffith, *39 Percent of Online Reviews Are Totally Unreliable,* PCMag.com (Nov. 7, 2019), *https://www.pcmag.com/news/371796/39-percent-of-online-reviews-are-totally-unreliable.*

[41] Bettie Cross, *Up to 30 percent of online reviews are fake and most consumers can't tell the difference,* CBS Austin, (Nov. 1, 2022), *https://cbsaustin.com/news/local/up-to-30-of-online-reviews-are-fake-and-most-consumers-cant-tell-the-difference.*

[42] Bart de Langhe *et al, Navigating by the Stars: Investigating the Actual and Perceived Validity of Online User Ratings,* Journal of Consumer Research, Volume 42, Issue 6 at 818–19 (April 2016) *https://www.colorado.edu/business/sites/default/files/attached-files/jcr_2016_de_langhe_fernbach_lichtenstein_0.pdf; see also* Jake Swearingen, Hijacked Reviews on Amazon Can Trick Shoppers, Consumer Reports (Aug. 26, 2019), *https://www.consumerreports.org/customer-reviews-ratings/hijacked-reviews-on-amazon-can-trick-shoppers/.*

[43] In the Matter of Rytr, LLC, Fed. Trade Comm'n, File No. 232–3052, Complaint, (Sep. 25, 2024), *https://advocacy.consumerreports.org/wp-content/uploads/2022/09/CR-Endorsement-Guides-comments-September-2022-3.pdf;* Consumer Reports filed a comment on the Rytr proceeding in support of the settlement, arguing it was in the public interest and that Rytr's product could only be reasonably used for fraudulent purposes. *See* Justin Brookman *et al.,* Consumer Reports files comment in support of FTC's settlement with Rytr, (Nov. 4, 2024), *https://advocacy.consumerreports.org/research/consumer-reports-files-comment-in-support-of-ftcs-settlement-with-rytr/.*

The role of tools and platforms

In the vast majority of cases, scammers and fraudsters do not create AI tools themselves—instead they take advantage of commercially available resources (many of which are free or at least very low cost). Sometimes these are general purpose tools, such as ChatGPT. In other cases, they are specialized products, including products like Rytr that are overwhelmingly likely to be used predominantly for illegitimate purposes—such as pornographic deepfake generation and voice impersonation. Many of the purveyors of these high-risk applications fail to take basic precautions to try to deter bad actors from using their products for illegitimate purposes.[44]

Once created, scammers use other general purpose platforms to host their solicitations. Amazon is rife with fake reviews, and social media sites like YouTube and Facebook host the deepfake endorsement scams described in the previous section. Using Google to search for “voice impersonation tools” yields several different options for people to impersonate others’ voices, including several sponsored results.

Nearly all online tools and platforms engage in some degree of content moderation to root out illegal activity, but in general they are underincentivized to expend sufficient resources to protect consumers.[45] Platforms that host illegal content are often explicitly immunized from responsibility by Section 230 of the Communications Decency Act.[46] In many cases, they benefit directly from the fraud, whether because they are paid by fraudsters who use their tools, they derive advertising revenue from hosting fraudulent content, they derive commissions from fraudulently endorsed products, or they benefit indirectly through artificially augmented engagement metrics which drive investors.[47]

II. Existing Protections

Scams and fraud are already illegal under a variety of Federal and state civil and criminal laws. The Federal Trade Commission along with other regulators and prosecutors around the country bring numerous enforcement actions every year.[48] In September of this year, the FTC announced a law enforcement sweep entitled “Operation AI Comply” to take action against companies that have used AI to perpetrate fraud.[49] The Rytr enforcement action discussed earlier was part of that sweep, as well as a case against a company that overstated the capabilities of their AI tool, and cases against companies that promoted fraudulent AI-powered business opportunities.

We are also seeing policymakers around the country enacting new laws to try to address potential abuses of AI. More than half the states have enacted laws prohibiting the use of AI to generate nonconsensual pornographic images.[50] Some states have expanded existing right-of-publicity laws to forbid the creation of digital replicas of real persons without their permission (or in the case of the deceased, their

[44] Janus Rose, *AI Tools Make It Easy to Clone Someone's Voice Without Consent,* Proof, (Jun. 25, 2024), *https://www.proofnews.org/ai-tools-make-it-easy-to-clone-someones-voice-without-consent/.*

[45] Testimony of Laurel Lehman, Policy Analyst, Consumer Reports Before the United States House of Representatives Committee on Energy & Commerce Subcommittee on Consumer Protection and Commerce on “Holding Big Tech Accountable: Legislation To Protect Online Users,” (Mar. 1, 2022), *https://www.congress.gov/117/meeting/house/114439/witnesses/HHRG-117-IF17-Wstate-LehmanL-20220301.pdf.*

[46] 47 U.S. Code § 230, *https://www.law.cornell.edu/uscode/text/47/230.*

[47] Mark Scott, *Report: Social Media Networks Fail to Root Out Fake Accounts,* Politico (Dec. 6, 2019), *https://www.politico.com/news/2019/12/06/social-media-networks-fake-accounts-report-076939;* Nicholas Confessore *et al., The Follower Factory,* N.Y. Times (Jan. 27, 2018), *https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html.*

[48] *E.g.,* Press Release, *FTC Takes Action to Stop Online Business Opportunity Scam That Has Cost Consumers Millions,* Federal Trade Commission, (Oct. 28, 2024), *https://www.ftc.gov/news-events/news/press-releases/2024/10/ftc-takes-action-stop-online-business-opportunity-scam-has-cost-consumers-millions.*

[49] Press Release, *FTC Announces Crackdown on Deceptive AI Claims and Schemes,* Federal Trade Commission, (Sep. 25, 2024), *https://www.ftc.gov/news-events/news/press-releases/2024/09/ftc-announces-crackdown-deceptive-ai-claims-schemes.*

[50] Vittoria Elliott, *The U.S. Needs Deepfake Porn Laws. These States Are Leading the Way,* Wired, (Sep. 5, 2024), *https://www.wired.com/story/deepfake-ai-porn-laws/; Most States Have Enacted Sexual Deepfake Laws,* multistate.ai, (Jun. 28, 2024), *https://www.multistate.ai/updates/vol-32.*

estates).[51] Twenty states have enacted comprehensive privacy laws since 2018,[52] and California passed the DELETE Act last year to make it easier for consumers to erase data broker records which could be used for targeted scams.[53] Colorado passed the first comprehensive bill designed to address potential bias in AI systems used in high-stakes decisions; the bill also requires consumer-facing AI systems to be labeled.[54] The state of California has initiated rulemaking proceedings under its privacy statute to enact similar protections for decision-making AI.[55] California also enacted an AI transparency law designed to address deceptive deepfakes: it requires generative AI products to offer deepfake detection tools and to embed invisible latent identifiers in artificial content to reflect the provenance of the image.[56] In general, after decades of failure to update the law to address the threats posed by new technologies such as the Internet and social media, state legislatures have been quicker to respond to some of the threats posed by artificial intelligence. The 2025 state legislative season is likely to see additional new laws related to AI enacted.[57]

Finally, regulators and others are ramping up user education efforts to warn consumers about the potential of AI scams and other AI-enabled fraud. Consumer Reports offers a free product called "Security Planner" to give people custom advice on the threats they are most concerned about;[58] Security Planner includes resources, for example, on how to spot phishing attempts and malicious websites posing as legitimate businesses.[59] We also publish and regularly update "The Consumer Reports Scam Protection Guide" that contains the latest information about evolving tactics.[60] Others like the FTC,[61] New York City,[62] and the Electronic Frontier Foundation[63] offer similar materials. The Public Interest Research Group offers helpful advice on how consumers can spot potential fake reviews.[64]

Consumer Reports is also part of a broader consumer awareness campaign called "Take 9" designed to train American consumers to pause and "take nine" seconds to consider calls to action that may be scams.[65] The campaign was launched earlier this fall and to date has generated 214 million impressions, 152,000 engagements, and a 14 percent engagement rate. Sentiment analysis of online mentions about the campaign highlights a clear trend: neutral and positive sentiments are growing at a significantly higher rate. In contrast, negative sentiment remains low, showcasing the campaign's initial resonance and positive impact.

Finally, developers are working on new tools—often themselves powered by AI—to identify artificial content. Mozilla offers a browser extension called "Fakespot" designed to identify potential fraudulent reviews when consumers are shopping online.[66] Polyguard offers a voice communication app for wealth managers (likely tar-

[51] CA AB–2602 Contracts against public policy: personal or professional services: digital replicas (2024),*https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240AB2602;* CA AB–1836 Use of likeness: digital replica (2024), *https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240AB1836;* TN HB2091 "Ensuring Likeness, Voice, and Image Security (ELVIS) Act of 2024," *https://legiscan.com/TN/text/HB2091/id/2900923*.

[52] *Which States Have Consumer Data Privacy Laws?,* Bloomberg Law, (Sep. 10, 2024), *https://pro.bloomberglaw.com/insights/privacy/state-privacy-legislation-tracker/*.

[53] SB–362 Data broker registration: accessible deletion mechanism (2023), *https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240SB362.*

[54] CO Senate Bill 24–205, The Colorado AI Act (2024), *https://leg.colorado.gov/sites/default/files/2024a_205_signed.pdf.*

[55] Press Release, *CPPA Adopts New Regulations for Data Brokers and Advances ADMT Rulemaking Package,* California Privacy Protection Agency, (Nov. 8, 2024), *https://cppa.ca.gov/announcements/2024/20241108_2.html.*

[56] CA SB–942 California AI Transparency Act (2024), *https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB942.*

[57] Jacob Rubenstein, *CR AI Fellow: Building a State Level AI Policy Tracker,* Consumer Reports Innovation Blog, (Sep. 12, 2024), *https://innovation.consumerreports.org/cr-ai-fellow-building-a-state-level-ai-policy-tracker/*.

[58] Security Planner, Consumer Reports, *https://securityplanner.consumerreports.org/*.

[59] *Spot Malicious Sites and Phishing Attempts,* Consumer Reports Security Planner, *https://securityplanner.consumerreports.org/tool/protect-yourself-from-phishing.*

[60] Janet Siroti, *The Consumer Reports Scam Protection Guide,* Consumer Reports, (Jul. 6, 2023), *https://www.consumerreports.org/money/scams-fraud/how-to-protect-yourself-from-scams-and-fraud-a6839928990/*.

[61] *Scams,* Federal Trade Commission, *https://consumer.ftc.gov/scams.*

[62] Tips on AI-Related Scams, NYC.gov, *https://www.nyc.gov/site/dca/consumers/artificial-intelligence-scam-tips.page.*

[63] *How to: Avoid Phishing Attacks,* Electronic Frontier Foundation Surveillance Self-Defense, (Jun. 24, 2024), *https://ssd.eff.org/module/how-avoid-phishing-attacks.*

[64] *How to recognize fake online reviews of products and services,* U.S. PIRG Education Fund, (Mar. 10, 2022), *https://pirg.org/edfund/resources/how-to-recognize-fake-online-reviews-of-products-and-services/*.

[65] *Nine Seconds for a Safer World,* Take9, *https://pausetake9.org/*.

[66] *Use AI to detect fake reviews and scams,* Fakespot, *https://www.fakespot.com/*.

gets for voice impersonation schemes) to identify potential calls from AI generated voice clones.[67] TrueMedia.org is a nonprofit organization dedicated to helping identify synthetic deepfake content.[68] As discussed above, recently enacted California legislation will also mandate that generative AI platforms create and offer AI deepfake detection tools which will hopefully increase the sophistication and adoption of such tools.[69]

III. Solutions

Responding to the new waves of fraud and scams powered by increasingly sophisticated AI is going to take a combination of legislation, enforcement, education, and cooperation from industry.

Stronger enforcement bodies

Fraud and scams are already illegal. However, because of insufficient enforcement—or consequences when caught—there is not enough deterrence against potential scammers. The FTC recently brought a handful of AI enforcement cases but those five actions are unlikely to meaningfully stem the already powerful wave of AI-power fraud.[70]

Currently, the FTC only has 1,292 FTEs total to pursue both its competition and consumer protection missions.[71] This number has been roughly flat over the past fourteen years, and actually represents a decrease from 1,746 FTEs in 1979. Put another way, since that time, the economy has grown nearly three times while the FTC's capacity has decreased by more than a quarter. The FTC is expected to hold giant sophisticated tech giants accountable for their transgressions, but they are severely hamstrung by unjustifiable resource constraints.

Even when the FTC does manage to bring a case, they often cannot get meaningful relief from the wrongdoer. For most violations of Section 5 of the FTC Act, the FTC cannot get statutory penalties from offenders. Historically, the FTC was at least able to obtain restitution—to get back the money that consumers lost to fraudsters. However, in 2021, the Supreme Court held that the FTC's enabling statute doesn't even give them that limited authority in many instances.[72] Despite bipartisan agreement that the FTC should be empowered to, at the very least, obtain the disgorgement of fraudulent gains from wrongdoers, Congress has failed to enact legislation to restore that power.[73]

Congress should grant the FTC additional resources to hire attorneys and technologists, and expand legal powers in order to allow the agency to keep pace with the threats that plague the modern economy.

Clearer platform accountability rules

Companies that offer AI tools and online platforms need clearer responsibility about how they respond to bad actors' use of those services. This could potentially be done using existing law. Section 5 of the Federal Trade Commission Act prohibits business practices that lead to significant consumer injury when that injury is not avoidable by consumers and the injury is not offset by countervailing benefits to consumers or competition.[74]

The FTC has long held that companies' failure to take action to identify and remediate harmful uses by bad actors of their products will in many cases be an unfair business practice. One analogous line of cases is the FTC's enforcement actions on

[67] *Trusted relationships demand trusted communications.,* Polyguard, *https://www.polyguard.ai/.*

[68] *Identifying Political Deepfakes in Social Media using AI,* TrueMedia.org, *https://www.truemedia.org/; see also* Cade Metz and Tiffany Hsu, *An A.I. Researcher Takes On Election Deepfakes,* New York Times, (Apr. 2, 2024), *https://www.nytimes.com/2024/04/02/technology/an-ai-researcher-takes-on-election-deepfakes.html.*

[69] CA SB–942 California AI Transparency Act (2024), *https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB942.*

[70] Press Release, *FTC Announces Crackdown on Deceptive AI Claims and Schemes,* Federal Trade Commission, (Sep. 25, 2024), *https://www.ftc.gov/news-events/news/press-releases/2024/09/ftc-announces-crackdown-deceptive-ai-claims-schemes.*

[71] *FTC Appropriation and Full-TIme Equivalent (FTE) History,* Federal Trade Commission, *https://www.ftc.gov/about-ftc/bureaus-offices/office-executive-director/financial-management-office/ftc-appropriation.*

[72] *AMG Capital Management, LLC v. Federal Trade Commission,*141 S. Ct. 1341 (2021), *https://www.supremecourt.gov/opinions/20pdf/19-508_l6gn.pdf.*

[73] Testimony of Anna Laitin, Director, Financial Fairness and Legislative Strategy, Consumer Reports Before the House of Representatives Committee on Energy & Commerce Subcommittee on Consumer Protection and Commerce on "The Consumer Protection and Recovery Act: Returning Money to Defrauded Consumers," (Apr. 27, 2021), *https://www.congress.gov/117/meeting/house/112501/witnesses/HHRG-117-IF17-Wstate-LaitinA-20210427.pdf.*

[74] 15 U.S. Code § 45, *https://www.law.cornell.edu/uscode/text/15/45.*

data security. In nearly a hundred cases since 2005, the FTC has said that companies have a legal obligation to anticipate and respond to ways that attackers could misuse their systems to gain access to consumers' personal information.[75] In these cases, the FTC has said that companies' failure to take steps to remediate likely abuses by third parties caused a substantial likelihood of injury that was unavoidable by consumers and not offset by countervailing benefits to consumers or competition. As just one example, earlier this year, the FTC brought an action against the security camera company Verkada for failure to take steps to prevent attackers from accessing video feeds from consumers' cameras.[76]

Beyond data security, the FTC has held companies responsible for how others use their products to cause harm to consumers.[77] For example, the FTC successfully sued QChex for violating Section 5 for allowing any customer to create checks for any bank account number without implementing reasonable safeguards to ensure that fraudsters were not creating checks for accounts they did not control. In that case, QChex's failure to take steps to prevent foreseeable harmful and illegal uses constituted an unfair business practice.

It is important to note that an obligation to identify and remediate likely harmful behaviors does not amount to strict liability for any harm caused by another bad actor using a company's product. Section 5's requirement that any harm not be offset by countervailing benefits to consumers or competition means that companies are not expected to spend unlimited resources to try to chase down potential offenders. Instead, the FTC only intervenes when companies fail to take cost-effective measures whose implementation would have prevented an even greater risk of injury.

Product design is also an important consideration in assessing the extent to which a company must take steps to remediate potential harm from bad faith actors. If the potential harms from a platform are especially significant, or the platform's design makes it likely that it will be used for harmful purposes, then companies should have a greater obligation to expend resources to remediate those uses. QChex, for example, allowed attackers to generate checks on consumers' bank accounts; given the high risk of substantial financial harm, the company had an obligation to ensure that the check writers in fact controlled those accounts and to monitor and respond to complaints of fraud. If a company creates a product that has a high likelihood of being used for illegitimate purposes, it should have a greater obligation to take steps to account for those harms to ensure the harms do not outweigh any potential benefits to consumers from the product.[78]

As such, generative AI products that are likely to be predominantly used for harm—such as Rytr's review generation service or voice impersonation companies—should have heightened obligations to address uses for illegal purposes (if they should be made commercially available at all). Those products are very likely to lead to significant consumer injury and consumers are unable to reasonably avoid

[75] See Press Release, *BJ's Wholesale Club Settles FTC Charges,* Federal Trade Commission, (Jun. 16, 2005), *https://www.ftc.gov/news-events/news/press-releases/2005/06/bjs-wholesale-club-settles-ftc-charges; Press Release, DSW Inc. Settles FTC Charges,* Federal Trade Commission, (Dec. 1, 2005), *https://www.ftc.gov/news-events/news/press-releases/2005/12/dsw-inc-settles-ftc-charges;* Press Release, *FTC Releases 2023 Privacy and Data Security Update,* Federal Trade Commission, (Mar. 28, 2024), *https://www.ftc.gov/news-events/news/press-releases/2024/03/ftc-releases-2023-privacy-data-security-up date;* Staff Report, *Start with Security: A Guide for Business,* Federal Trade Commission (Jul. 2017), *https://www.ftc.gov/system/files/documents/plain-language/pdf0205-startwithsecurity.pdf.*

[76] See Press Release, *FTC Takes Action Against Security Camera Firm Verkada over Charges it Failed to Secure Videos, Other Personal Data and Violated CAN–SPAM Act,* Federal Trade Commission, (Aug. 30, 2024), *https://www.ftc.gov/news-events/news/press-releases/2024/08/ftc-takes-action-against-security-camera-firm-verkada-over-charges-it-failed-secure-videos-other.*

[77] See, *e.g.,* Press Release, *FTC Sues Walmart for Facilitating Money Transfer Fraud That Fleeced Customers Out of Hundreds of Millions,* Federal Trade Commission, (Jun. 28, 2022), *https://www.ftc.gov/news-events/news/press-releases/2022/06/ftc-sues-walmart-facilitating-money-transfer-fraud-fleeced-customers-out-hundreds-millions;* Press Release, *U.S. Circuit Court Finds Operator of Affiliate Marketing Network Responsible for Deceptive Third-Party Claims Made for LeanSpa Weight-loss Supplement,* Federal Trade Commission, (Oct. 4, 2016), *https://www.ftc.gov/news-events/news/press-releases/2016/10/us-circuit-court-finds-operator-affiliate-marketing-network-responsible-deceptive-third-party-claims;* Press Release, *Court Orders Permanent Halt to Illegal Qchex Check Processing Operation Court Finds Qchex Unfair Practices Created a Dinner Bell for Fraudsters Operators to Give Up All Their Ill-Gotten Gains,* Federal Trade Commission, (Feb. 9, 2009), *https://www.ftc.gov/news-events/news/press-releases/2009/02/court-orders-permanent-halt-illegal-qchex-check-processing-operation-court-finds-qchex-unfair.*

[78] Press Release, *U.S. Circuit Court Finds Operator of Affiliate Marketing Network Responsible for Deceptive Third-Party Claims Made for LeanSpa Weight-loss Supplement,* Federal Trade Commission, (Oct. 4, 2016), *https://www.ftc.gov/news-events/news/press-releases/2016/10/us-circuit-court-finds-operator-affiliate-marketing-network-responsible-deceptive-third-party-claims.*

them—to the contrary, the services are designed to create content that is indistinguishable from authentic content. There are limited positive use cases for these tools, so the harms caused by making these platforms generally available without reasonable safeguards in place is unlikely to be outweighed by countervailing benefits.

For more general purpose tools, the calculus is significantly more complicated. ChatGPT, for example, is a multipurpose system designed to respond to any number of constantly changing prompts—the cost of anticipating and responding to every potential abuse of the system is substantially higher. In fact, the developers of ChatGPT do consider potential misuse by bad actors and do put some limits on how the platform can be used. For example, ChatGPT regularly updates and publishes a system card identifying “Key Areas of Risk Evaluation & Mitigation,” including “unauthorized voice generation” and “generating erotic and violent speech.”[79] Further, making changes to account for harmful uses could also potentially constrain known or unknown positive uses of ChatGPT—another potential countervailing benefit that is less likely for narrower tools designed for specific tasks.

Nevertheless, there is a strong case that the developers of general purpose generative AI products should take more aggressive measures to prevent or deter obvious abuses (as discussed above, general purpose services comply with requests to generate multiple “fake reviews”). The extent to which such a multipurpose platform should take steps to respond to different threats is a complex question, balancing the costs of potential harm with the costs of remediation and potential limitations of beneficial uses. The same goes for platforms that host fraudulent content; if their services are causing significant harm and there are cost-effective measures they could employ to remediate that harm, they should do so.

Clarifying tool and platform responsibility for customer abuse could be done through enforcement under Section 5 and comparable state consumer protection laws. Or new legal protections could be enacted to specify what steps these companies should take and under what circumstances when their products are used to defraud consumers.

Transparency obligations

As a general matter, consumers deserve to know whether the content they’re interacting with is real or AI-generated. Content creators and companies should be labeling AI-generated content and chatbots as such. The fact that content is AI-generated should be communicated prominently and contextually in such a way that an ordinary consumers is likely to notice, through visual labeling (or in the case of AI-generated phone calls, through an introductory statement, as the FCC recently proposed in a rulemaking on AI-generated robocalls).[80] It should also be communicated latently through standardized metadata, watermarks, or other technology to allow platforms and agents to automatically identify content as synthetic—this approach was recently mandated in California in legislation enacted in September of this year.[81]

However, mandated transparency has limitations too, as bad actors will simply fail to provide prominent disclosures, and will endeavor to strip our latent identifiers imposed by generative platforms—or they will turn to smaller, malicious, or open-source platforms that are not covered by or otherwise do not comply with transparency obligations. For this reason, some advocates have argued against transparency obligations, noting that adversarial transparency obligations have historically been ineffective.[82] Nevertheless, we think there is a benefit in requiring transparency from actors who may be deterred from breaking the law, and over time

[79] GPT–4o System Card, OpenAI, (Aug. 8, 2024), *https://openai.com/index/gpt-4o-system-card/.*

[80] Comments of Consumer Reports on Implications on Artificial Intelligence Technologies on Protecting Consumers From Unwanted Robocalls and Robotexts, CG Docket No. 23–362, (Oct. 10, 2024), *https://advocacy.consumerreports.org/wp-content/uploads/2024/10/CR-Comment-on-FCC–AI-Robocall-Rulemaking-.pdf.* 50,000 consumers signed onto our petition in support of our comments calling for transparency in AI-generated robocalls. *See* Grace Gedye, 50,000 consumers support FCC AI robocall rulemaking, Consumer Reports Advocacy, (Oct. 10, 2024), *https://advocacy.consumerreports.org/research/50000-consumers-support-fcc-ai-robocall-rulemaking/.*

[81] CA SB–942 California AI Transparency Act (2024), *https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB942.*

[82] Jacob Hoffman-Andrews, *AI Watermarking Won’t Curb Disinformation,* Electronic Frontier Foundation, (Jan. 5, 2024), *https://www.eff.org/deeplinks/2024/01/ai-watermarking-wont-curb-disinformation.*

content identification (including reliably authenticating when content is organic and legitimate) may improve.[83]

Stronger privacy and security laws

Scams are much more effective when attackers have access to personal information in order to customize a solicitation. However, the United States's privacy laws are weak, and hundreds of unregulated data brokers are able to amass detailed dossiers about all of us which are then sold to anyone willing to pay for them.[84]

The Federal government has no comprehensive privacy law, and instead only has a handful of laws of varying strength covering sensitive categories of personal information such as medical, financial, and kids' data. Over the past six years, twenty states have passed their own general privacy laws, though most of those laws are too weak to meaningfully stem the flow of personal information to data brokers and advertisers.[85]

Policymakers should enact stronger privacy rules based on the principle of *data minimization*—meaning companies should only be collecting, processing, sharing, and retaining data as is reasonably necessary to deliver the goods or services requested by consumers.[86] Such a model would protect personal data *by default* rather than subject consumers to relentless requests for "opt-in" consent for superfluous data usage or forcing consumers to navigate innumerable "opt-out" mechanisms.[87]

Citizen education and better tools

Finally, consumers have a role to play too, and digital citizens will have to become more savvy and discriminating in a world where even very realistic looking and sounding content may be entirely AI-generated. For the last three years, Consumer Reports has published a "Cyber Readiness Report" which draws on nationally representative surveys to track the adoption of cybersecurity best practices over time.[88] While a majority of respondents did exhibit awareness of the importance of unique passwords, software updates, and multifactor authentication, adoption of more sophisticated techniques (such as use of password managers and tracker blockers) is lagging; moreover, adoption of cybersecurity best practices in general has remained fairly flat over the past three years. Institutions will need to adapt to find ways to encourage consumers to adopt more sophisticated protections over time, including protections designed to protect consumers from AI-generated deepfake scams. At the same time, researchers in industry, academia, civil society, and government will have to continue to develop new tools to help consumers identify inauthentic content. Over time, these tools need to become seamlessly embedded into browsers, mobile phone operating systems, and other platforms that consumers use to access online content.

Thank you very much for the opportunity to testify today, and I look forward to answering your questions.

[83] Grace Gedye, *CR submits testimony AI and consumer protection to New York Assembly,* Consumer Reports Advocacy, (Sep. 25, 2024), *https://advocacy.consumerreports.org/research/cr-submits-testimony-ai-and-consumer-protection-to-new-york-assembly/.*

[84] This Committee has published a detailed investigation into data brokers, though at this point the report is over ten years old. *See* Office of Oversight and Investigations Majority Staff, A Review of the Data Broker Industry: Collection, Use, and Sale of Consumer Data for Marketing Purposes, Committee on Commerce, Science, and Transportation, (Dec. 18, 2013), *https://www.commerce.senate.gov/services/files/0d2b3642-6221-4888-a631-08f2f255b577.* The situation has not materially improved in the meantime.

[85] *The State of Privacy: How state "privacy" laws fail to protect privacy and what they can do better,* Electronic Privacy Information Center and U.S. PIRG Education Fund, (Feb. 2024), *https://publicinterestnetwork.org/wp-content/uploads/2024/01/State-of-Privacy-Feb.-2024.pdf.*

[86] *How the FTC Can Mandate Data Minimization Through a Section 5 Unfairness Rulemaking,* Consumer Reports and the Electronic Privacy Information Center, (Jan. 26, 2022), *https://advocacy.consumerreports.org/wp-content/uploads/2022/01/CR_Epic_FTCDataMinimization_012522_VF_.pdf.*

[87] In 2021, Consumer Reports published model privacy legislation based on the concept of data minimization. *See Model State Privacy Act,* Consumer Reports Advocacy, (Feb. 2021), *https://advocacy.consumerreports.org/wp-content/uploads/2021/02/CR_Model-State-Privacy-Act_022321_vf.pdf.* Earlier this year, Consumer Reports and the Electonic Privacy Information Center published a compromise approach based on the Connecticut privacy legislation that has served as a model for several other states. *See* Press Release, *Consumer Reports and the Electronic Privacy Information Center unveil new model legislation to protect the privacy of American consumers,* Consumer Reports Advocacy, (Sep. 24, 2024), *https://advocacy.consumerreports.org/press_release/consumer-reports-and-the-electronic-privacy-information-center-unveil-new-model-legislation-to-protect-the-privacy-of-american-consumers/.*

[88] Yael Grauer, *New Report: 2024 Consumer Cyber Readiness,* Consumer Reports Innovation Blog, (Oct. 1, 2024), at 4 *https://innovation.consumerreports.org/new-report-2024-consumer-cyber-readiness/.*

Senator HICKENLOOPER. Thank you, Mr. Brookman.

Next, Mr. Mounir Ibrahim, who is the Chief Communications Officer and Head of Public Affairs at Truepic, also a former State Department official, I think, so brings a broad wealth of kind of more international, more global experience perspective to this.

STATEMENT OF MOUNIR IBRAHIM, CHIEF COMMUNICATIONS OFFICER AND HEAD OF PUBLIC AFFAIRS, TRUEPIC

Mr. IBRAHIM. Thank you, Chairman Hickenlooper, Ranking Member Blackburn and members of this committee.

My name is Mounir Ibrahim. I am with Truepic, a technology company that is focused on digital content, authenticity, and transparency online today.

As you noted, Senator, I was a former Foreign Service Officer with the U.S. Department of State and it was there I saw firsthand the importance and critical need for transparency and authenticity in digital content from conflict zones around the world including on the ground in Damascus, Syria.

I fundamentally believe that deciphering what is human generated from AI is one of the most pressing challenges we have today. I applaud this committee for its leadership on this issue.

I would like to address the threat very briefly, echoing my colleagues here, to local communities and everyday individuals. This dynamic is most clearly seen in the prevalent and alarming rise of nonconsensual pornography as we heard, often targeting young women and even minors.

Meanwhile, catfishing and sextortion scams are on the rise, often powered by AI. I fear we are witnessing the early stages of AI being weaponized against local communities and individuals who do not have the resilience or resources to defend themselves.

We know that businesses face the same challenges, too. It is reported that deepfake phishing scams on businesses grew by 3,000 percent in the last year alone, 700 percent in the FinTech sector.

These trends harm consumers and jeopardize America's economic competitiveness in the AIH. The reality is there is no single solution to stop this problem. However, there are effective strategies to mitigate the threat and one key approach I would like to talk about is digital content provenance.

Content provenance is cryptographically and securely attaching information metadata to digital content, images, videos, audio that we see online every day. It can help you tell if something is AI generated or if it was camera captured and what edits might have been made along the way.

Today, content credentials are being used to protect businesses and people alike. It is being led by the Coalition for Content Provenance and Authenticity, the C2PA, of which Truepic is a proud steering committee member. Very briefly, I would like to share how Truepic is leveraging content credentials with its partners in business and for consumer benefit.

Business credentialing and recredentialing is one of the fastest growing industries that is adopting content provenance. We work with partners like Equifax, TransUnion, Dunn & Bradstreet, all of which to securely transform their digital verification process backed by provenance.

This protects consumers by enabling our partners to verify the authenticity of those people buying credit reports, thereby helping safeguard data.

Other industries like insurance, product recalls, and others are all embracing the same approach. Moving on to more consumer-facing benefits, OpenAI's ChatGPT strengthen safeguards for misuse.

We are honored that our technology helps back that with a certificate authority that allows these images that come from ChatGPT to be read and deciphered by adoptive social media platforms like LinkedIn, et cetera.

Last month we produced the world's first authenticated video with content credentials on YouTube and thereby you can see the content credentials, seeing that it is in fact captured with a camera.

We are working with partners like Qualcomm to deliver these benefits directly on chip sets so that the future smart phones we all own and operate that will both create AI and capture authentic content can add these content credentials if chosen so.

While these are all highly encouraging points, there are challenges and I would like to be transparent about those challenges. The first challenge is adoption for content provenance to truly transform and create the—make the Internet a more authentic and transparent place. We need adoption. We expect the C2PA standard to become a global standard under the ISO next year but more must be done.

Education—even if all our digital content has content credentials it is imperative that content consumers and businesses and those interacting with it know what they mean and what they do not mean. These are not blind stamps of trust and that is a critical piece.

Moving forward, I think Congress can help create this—help make this a reality. First and foremost, education and funding. There are countless academic institutions and research institutions that are looking at how content provenance can scale and can improve the health of the internet, and Congress can help support them.

Engagement—hearings like this to help emphasize that content provenance is not just a safeguard but it also is an unlocker for innovation and efficiency across government and the private sector.

Thank you very much, and I look forward to your questions.

[The prepared statement of Mr. Ibrahim follows:]

PREPARED STATEMENT OF MOUNIR IBRAHIM, CHIEF COMMUNICATIONS OFFICER AND HEAD OF PUBLIC AFFAIRS, TRUEPIC

Thank you, Chairman Hickenlooper, Ranking Member Blackburn, and esteemed members of this subcommittee. My name is Mounir Ibrahim, and I represent Truepic—a technology company dedicated to providing essential transparency and authenticity tools for the digital content we encounter daily.

The importance of understanding the origins and authenticity of digital content was deeply impressed upon me during my time as a Foreign Service Officer with the U.S. Department of State, where I worked on various global conflicts, including serving on the ground at our Embassy in Damascus, Syria. My experiences as a diplomat exposed me to the urgency of verifying the authenticity of digital content at the highest levels of national and global security decision-making.

I fundamentally believe deciphering what is human-created from AI will become one of the most pressing challenges across all aspects of life.

At Truepic, we recognize that our increasingly digitized lives rely on digital content for decisions—personal, business, and governmental. From insurance claims and banking audits to social media feeds, online profiles, and even images from conflict zones—digital content shapes the decisions we make every day.

I applaud the subcommittee's focus on the impact of digital content on people and consumers in today's AI age. I also thank you for your continued leadership in supporting transparency online.

Threat Landscape

You are undoubtedly aware of the rapid growth of AI platforms, many of which are publicly available, open-sourced, and easy to access. While much attention has rightly been given to how synthetic media and deepfakes could impact elections, mischaracterize leaders, and manipulate markets, I would like to briefly address *the threat to local communities* and everyday individuals who lack the resources to quickly and easily debunk false content.

Local schools, community leaders, businesses, and law enforcement face significant challenges in distinguishing human-created from synthetic content online, leaving them vulnerable as bad actors exploit easily accessible AI tools with little or no technical expertise.

This dynamic is most clearly seen in the prevalence and alarming rise of non-consensual pornography, often targeting young women, even minors. A *recent study* across 10 countries performed by researchers in Australia and the United States found that 2.2 percent of respondents were victims of deepfake pornography, while 1.8 percent admitted to creating or sharing it. Meanwhile, Catphishing and *Sextortion* scams are rapidly increasing and more often powered by AI technology.

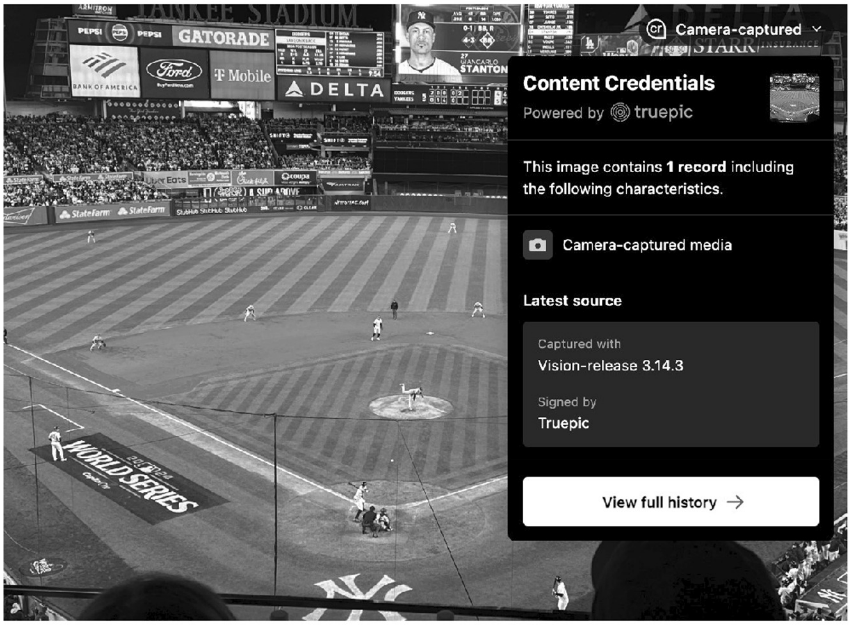
I fear we are witnessing the early stages of AI being weaponized against local communities and individuals who lack the resources to defend themselves. AI-driven visual deception is rapidly expanding beyond non-consensual pornography into local politics, schools, and business fraud. In *New York* and *Maryland,* for example, deepfake videos falsely depicted school leaders making offensive and racist remarks, uprooting communities and making national news. In Louisiana, *bad actors on platforms* like 4Chan have shared tools to create deepfakes targeting judges, prosecutors, and defendants during locally streamed parole board hearings.

We know that businesses also face the same challenges. It has been reported that *deepfake phishing scams grew by 3000 percent* and deepfake incidents in the fintech sector *increased by 700 percent* last year alone. These trends harm consumers, small business owners, and jeopardize America's economic competitiveness.

What Can Be Done?

The reality is that there is no single solution to halt the growing trend of visual deception. However, there are effective strategies to mitigate the threat, elevate human-created content, and enhance transparency across the internet. I will defer to my esteemed colleague on the panel to discuss the capabilities of the detection of AI content, but one key approach I want to emphasize to this committee is the importance and potential of digital content provenance.

Content provenance securely and cryptographically attaches information (metadata), known as *Content Credentials,* to photos, videos, or audio so consumers can understand where it came from, how it was created, and whether it has been edited. For example, Content Credentials can show if something was made by a camera or generated by AI, when and where it was created, and any changes it has gone through. This helps people understand what they see online by providing a clear record of its origins and history.



Content Credentials

This image contains **1 record** including the following characteristics.

Camera-captured media

Latest source

Captured with
Vision-release 3.14.3

Signed by
Truepic

Image of 2024 World Series with Content Credentials

This is not hyperbole. Today, Content Credentials are being used by many companies and leveraged on some of the world's most-used social media sites (LinkedIn, YouTube, Instagram, Facebook, etc.). The approach is driven by an open standard developed by the *Coalition for Content Provenance and Authenticity* (C2PA), of which Truepic is a proud steering committee member, along with Microsoft, Adobe, Google, OpenAI, BBC, and many others.

How Is It Being Used?

I would like to share how Truepic is leveraging Content Credentials and how that is helpful to consumers. First, let me begin with how the private industry has deployed Content Credentials as a necessary pre-requisite to operating in the AI world.

- Business credentialing and re-credentialing are quickly becoming one of the leading industries to embrace content provenance for verifying buyers of credit reports. Without verifiable and transparent content, companies face significant risks and become more vulnerable to AI-driven fraud. Provenance technology enables our partners like *Equifax,* Transunion, and *Dun & Bradstreet* to securely transform the verification process.
  - Content Credentials protect consumers by enabling our partners to digitally verify the authenticity of organizations purchasing credit reports. This helps ensure that potentially sensitive credit information is shared only with authorized and legitimate businesses, helping safeguard consumers from AI fraud and deception.
- Insurance is another critical industry that has digitized its processes and is subject to significant fraud that raises rates and complicates processes for consumers. With partners such as *EXL Service, USAA, Jewelers Mutual,* and many more, our partners are leveraging content provenance to streamline claims processing, underwriting, reduce fraud, and improve customer experiences. Other partners, like *Palomar Specialty Insurance,* deploy content provenance to address natural disasters, speeding up the process for and supporting victims through authenticated content.
  - Content Credentials protect consumers in insurance by ensuring claims and underwriting are supported by genuine and verifiable evidence, preventing

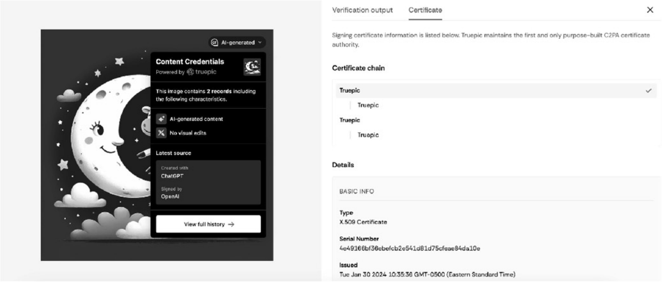
fraud that can inflate premiums for everyone. This benefits policyholders by allowing them to receive fair and timely resolutions.

We are excited to extend this impact to other consumer-focused areas like product recalls through partnerships with organizations like Sedgwick. Verifiable information through Content Credentials ensures that owners and victims of defective products can quickly prove possession and malfunction, allowing them to be compensated more efficiently and without unnecessary red tape, enhancing consumer protection.

Public Facing Benefit

Let me address how Content Credentials are being deployed to help power a more authentic and transparent online experience essential to help stem the deception of people online.

OpenAI's ChatGPT, including DALL&middot;E, is one of the most prominent AI text-to-image generation platforms. It has implemented stronger safeguards to combat misuse and prevent deception, an effort we are proud to support. Truepic plays a key role in this initiative by providing the Certificate Authority that backs the C2PA signature on AI-generated images, ensuring transparency and accountability. Content Credentials on OpenAI outputs help consumers because the Truepic certificate will be recognized by major social media platforms like LinkedIn, Instagram, or Facebook, and the consumer will be alerted the image is AI-created.



Content Credentials
Powered by truepic

This image contains **2 records** including the following characteristics.

AI-generated content
No visual edits

Latest source
Created with
ChatGPT
Signed by
OpenAI

Signing certificate information is listed below. Truepic maintains the first and only purpose-built C2PA certificate authority.

Certificate chain

Truepic ✓
Truepic
Truepic
Truepic

Details

BASIC INFO

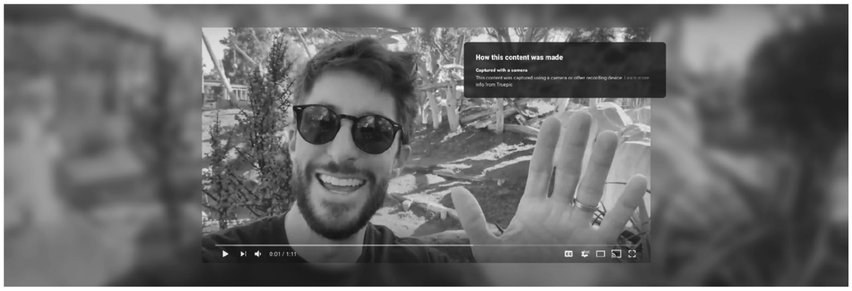
Type
X.509 Certificate

Serial Number
4e49166bf36ebefcb2e541d81d75cfeae84da10e

Issued
Tue Jan 30 2024 10:35:36 GMT-0500 (Eastern Standard Time)

Image created on ChatGPT with C2PA Content Credentials backed by Truepic's Certificate Authority

We are also working to help power any organization that wants to capture and share authentic video on YouTube. Last month, we produced the first human-created video with Content Credentials on YouTube, which much like the prior example, allows consumers to understand the origins of the video they are watching. This capability can be extended to any application that captures videos and places them on YouTube.



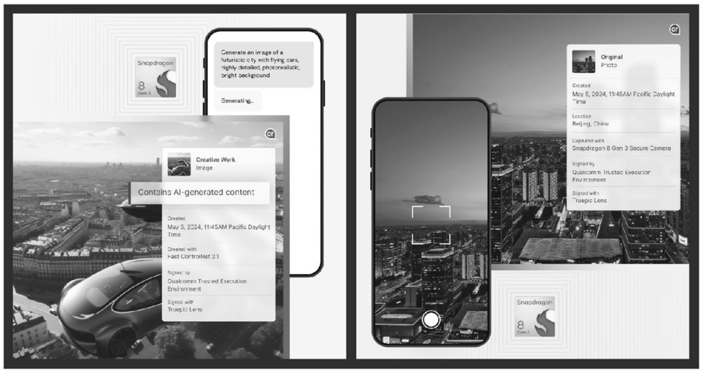
How this content was made

Captured with a camera

This content was captured using a camera or other recording device. Learn more info from Truepic

YouTube displays Content Credentials on authentically-produced videos

Another exciting example of this technology creating transparency for consumers lies in our smartphones. With AI increasingly accessible on mobile devices, we're proud to partner with Qualcomm to embed Truepic technology directly into Snapdragon chipsets. This enables on-device AI content to be tagged with Content Credentials. At the same time, creators can also choose to tag authentic images and videos, ensuring transparency in digital content while safeguarding privacy.



Generate an image of a futuristic city with flying cars, highly detailed, photorealistic, bright background

Generating...

Creative Work
Image

Contains AI-generated content

Created
May 5, 2024, 11:45AM Pacific Daylight Time

Created with
Fast ControlNet 2.1

Signed by
Qualcomm Trusted Execution Environment

Signed with
Truepic Lens

Original
Photo

Created
May 5, 2024, 11:45AM Pacific Daylight Time

Location
Beijing, China

Captured with
Snapdragon 8 Gen 3 Secure Camera

Signed by
Qualcomm Trusted Execution Environment

Signed with
Truepic Lens

Qualcomm and Truepic developed the first chipset to power digital content transparency across smartphones worldwide

Challenges

While we are highly encouraged by this progress, it is important to acknowledge that significant challenges persist.

- *Adoption:* For the open specification to reach maximum efficacy, we need the various components of the Internet to adopt and implement, this includes browsers, platforms, CDNs (Content Delivery Networks), and OEMs (Original Equipment Manufacturers). Without widespread adoption, Content Credentials can be easily lost when content moves to non-compliant platforms. The C2PA expects the International Standards Organization (ISO) to approve the specification as an open standard in the coming months, and we anticipate there will be further adoption; however, much more needs to be done.
- *Education:* We also need to work together—industry, government, and civil society to educate all stakeholders on what Content Credentials mean and, perhaps more importantly, do not mean. They are not meant and should not be considered stamps of blind trust. Rather, they are indicators that more information on that piece of content is available so that consumers can make more accurate decisions based on digital content. We believe this is essential and the C2PA, supported by a generous grant from Microsoft and OpenAI, has been working hard to accelerate educational efforts on the specification and approach.

Moving Forward & Recommendations

We are dedicated to working with partners and the C2PA community to ensure Content Credentials empower more informed decisions online. We also believe that government can be critical in advancing a more transparent internet, especially with the following:

- *Education & Funding:* We believe this is the most immediate area in which the government can support various education initiatives and research institutions examining how provenance and Content Credentials can be most effective in supporting consumers.
- *Engagement:* Hearings like this with policymakers, industry leaders, and civil society are essential to raising awareness. I would encourage looking at provenance and transparency beyond just a safety measure, and more importantly, as an ***opportunity*** that unlocks significant government and private sector efficiencies.
- *Existing Recommendations & Research:* In line with the bipartisan AI Insight Roadmap, we encourage this committee to consider how establishing provenance of digital content, for both synthetic and non-synthetic content, can be beneficial. NIST is also performing a critical role as it looks at how transparency in AI can be more readily available.

Transparency is essential as digitization accelerates and the line between human-created and synthetic content blurs. Without it, consumers face greater deception, businesses endure rising fraud, and governments risk miscalculation and exploitation. These challenges are significant, but with collaboration and the awareness driven by hearings like this, I am confident we can build a more transparent and authentic digital future.

Thank you for considering this testimony.

Senator HICKENLOOPER. Thank you very much, Mr. Ibrahim.

Next, Dorota Mani, who is an entrepreneur in her own right and has experienced some of the consequences of these fraudsters in a powerful way, and I really appreciate you being here.

STATEMENT OF DOROTA MANI, MOTHER OF DEEPFAKE PORNOGRAPHY VICTIM

Ms. MANI. Thank you for having me. Thank you for the opportunity to testify before this committee today. My name is Dorota Mani and I stand before you not only as an educator and advocate but also as a mother.

My 14-year-old daughter, along with her sophomore classmates at Westfield High School, was a confirmed victim of AI deepfake misuse perpetrated by her peers last year.

Boys in my daughter's grade used AI to generate sexually explicit images of her and other girls. AI technology, as complex and powerful as it is, presents both advancements and challenges to our society.

While AI can greatly enhance many aspects of our life it also has the potential to harm your constituents deeply. I am here not to dwell on how this incident has made us feel but, rather, to shift the narrative away from the victims and toward the perpetrators, the lack of accountability, lack of laws, and the urgency of implementing effective safeguards.

Based on personal experience, I strongly believe there is a critical missing component in our approach to artificial intelligence, which is education to prevent misuse.

While not an area directly related to this committee, I do believe school districts should consider implementing AI literacy programs to ensure our children comprehend the implications and responsibilities associated with using such powerful technologies safely and ethically.

But more than that, those programs will prepare our children not only to protect themselves but also to thrive in increasingly digital world.

Unfortunately, AI and its capabilities will allow them to critically assess and navigate—understanding AI and its capabilities will allow them to critically assess and navigate potential manipulations by AI across several aspects of life.

They will learn to cross reference and validate what they see, read, and hear, therefore avoiding manipulation and scams. At the same time, they can harness this tool to bridge gaps between affluent and underserved communities as well as rural and urban schools, in this manner providing equal access to resources and opportunities.

Simultaneously, we need robust regulations and legislation for the specific harms AI creates like deepfake sexually explicit images.

Currently, the absence of school policies on AI, alongside a lack of specific civil and criminal laws, leaves a significant gap in our ability to hold bad actors accountable.

In the incident involving my daughter, the lack of AI specific guidelines at the school led to minimal disciplinary actions against the perpetrators who not only remained at the school and continued to attend the classes but also represent the school till this day in sports teams.

The repeated excuse by the school was the absence of state and Federal laws. Even though there are many important AI bills like the Shield Defiance Act, Labeling Act, just to name a few, Senator Cruz's Take it Down Act deeply resonates with my daughter and I as it allows the victims to take control over their image by ensuring that those images can be easily taken off any sites where they can be seen by anyone, protecting the victim's reputation.

It is also—it also creates a Federal law that criminalizes the publication of deepfakes, giving schools a tool to respond against AI misconduct instead of hiding behind the lack of laws.

I thank the members of this committee for unanimously passing this legislation in July and I call for members of the Senate to pass it urgently to protect my daughters and other girls in this country.

We should also consider further AI regulations that protects people from deepfake images like a labeling tool for AI-generated content to inform recipients of its source. Deepfake images circulating within the digital ecosystem can harm victims professionally, educationally, personally, and emotionally and more, potentially destroying reputations and futures.

We should not try to reinvent the wheel but rather learn from existing models. I also would like to emphasize the urgent need to reform Section 230 especially given the rapid evolution of technology and the challenges posed by AI.

Just as laws have adapted for self-driving versus traditional vehicles it is crucial that Section 230 evolves to stay relevant and effective in today's digital landscape. Let us not misunderstood C2—that should not be touched—of Section 230.

To tackle the spread of harmful online content we should focus on solutions starting with a round table involving leaders from major content holding platforms like Microsoft, Google, Apple, and Amazon.

Those companies have the expertise and resources to enact significant changes immediately driven by social and ethical responsibility while legislators are crafting laws to hold bad actors accountable.

Thank you.

[The prepared statement of Ms. Mani follows:]

Prepared Statement of Dorota Mani, Mother of Deepfake Pornography Victim

Thank you for the opportunity to testify before the committee today. My name is Dorota Mani and I stand before you not only as an educator and advocate but also as a mother. My 14-year-old daughter, along with her sophomore classmates at Westfield High School, was a confirmed victim of AI deep fake misuse perpetrated by her peers last year. Boys in my daughter's grade used AI to generate sexually explicit images of her and other girls. AI technology, as complex and powerful as it is, presents both advancements and challenges to our society. While AI can greatly enhance many aspects of our lives, it also has the potential to harm your constituents deeply.

I am here not to dwell on how this incident has made us feel but rather to shift the narrative away from the victims and towards the perpetrators, the lack of accountability, lack of laws and the urgency of implementing effective safeguards.

Based on personal experience, I strongly believe there is a critical missing component in our approach to artificial intelligence, which is Education to Prevent Misuse. While not an area directly related to this Committee, I do believe school districts should consider implementing AI literacy programs, to ensure our children comprehend the implications and responsibilities associated with using such powerful technologies safely and ethically. But more than that, these programs will prepare our children not only to protect themselves but also to thrive in an increasingly digital world. Understanding AI and its capabilities will allow them to critically assess and navigate potential manipulations by AI across several aspects of life. They will learn to cross-reference and validate what they see, read, and hear, therefore avoiding manipulation, scams, or misinformation. At the same time, they can harness this tool to bridge gaps between affluent and underserved communities, as well as rural and urban schools, in this manner providing equal access to resources and opportunities.

Simultaneously, we need robust Regulation and Legislation for the specific harms AI creates, like deepfake sexually explicit images. Currently, the absence of explicit school policies on AI, alongside a lack of specific civil and criminal laws, leaves a significant gap in our ability to hold bad actors accountable. In the incident involving my daughter, the lack of AI-specific guidelines at her school led to minimal disciplinary action against the perpetrators, who not only remained at the school but continued to attend classes and represent the school in sports till this day. The repeated excuse by the school was the absence of state and Federal laws.

Even though there are many important AI bills, Senator Cruz's TAKE IT DOWN Act deeply resonates with my daughter and I as it allows the victims to take control over their image, by ensuring that these images can be easily taken off any site where they can be seen by anyone, protecting the victim's reputation. It also creates a Federal law that criminalizes the publication of deepfakes giving schools a tool to respond against AI misconduct, instead of hiding behind lack of laws. I thank the members of the Committee for unanimously passing the legislation in July and I call for members of the Senate to pass it urgently.

We should also consider further AI regulation that protects people from deepfake images, like a labeling tool for AI-generated content to inform recipients of its source. Deep fake images circulating within the digital ecosystem can harm victims professionally, educationally, personally, emotionally, and more, potentially destroying reputations and futures. We should not try to reinvent the wheel but rather learn from existing models

I also would like to emphasize the urgent need to reform Section 230, especially given the rapid evolution of technology and the challenges posed by AI. Just as laws have adapted for self-driving versus traditional vehicles, it's crucial that Section 230 evolves to stay relevant and effective in today's digital landscape.

To tackle the spread of harmful online content, we should focus on solutions, starting with a roundtable involving leaders from major content holding platforms like Microsoft, Google, Apple, and Amazon. These companies have the expertise and resources to enact significant changes immediately, driven by social and ethical responsibility, while legislators are crafting laws to hold bad actors accountable.

Similarly, financial giants such as AMEX, PayPal, Chase, and Visa play a fundamental role in our digital ecosystem. It's essential they too engage proactively to prevent their platforms from facilitating harmful activities. Collaborative efforts with these institutions can help establish a secure and ethical transactional environment, showcasing a commitment to responsible business practices.

To protect our children and ensure a safe learning environment, we must do more than just react, we need proactive education, stern policies, timely enforcement, and tough legislation that addresses the harms real people face.

As AI continues to filter through every aspect of our lives, from simple tasks to complex decision-making processes, it is imperative that we educate our society about its potential and dangers. AI, the new calculator, is here to stay. Let us face this challenge head-on, recognize the need for sector specific regulation directed at particular harms, and harness this tool to bridge educational and workforce gaps, providing opportunities to advance, regardless of background.

Thank you and I look forward to your questions.

Senator HICKENLOOPER. Thank you. Thank you each for being here.

Now, we will stop and there—this is a kind of a crazy day so I think people are going to be coming in and out but you will be on a nonfraudulent video record. So we will make sure that you will have—despite the lack of senators sitting right at the table we will make sure many of them see this. I know there is tremendous interest.

Mr. Brookman, why do I not start with you? We have seen AI technology evolve so rapidly the last few years I think we have to anticipate that that trend is going to continue and probably accelerate.

In previous hearings we have highlighted how comprehensive privacy legislation is an essential building block to addressing some of the harms presented by AI, those that are anticipated and those that have already happened.

As AI continues to evolve how would you see data privacy protections helping to mitigate fraud against consumers?

Mr. BROOKMAN. Yes. So a lot of these scams are social engineering attacks and those are a lot more effective when they have data about you, and in the current environment there is a lot of data out there available about us. There is, like, hundreds of data brokers. California has a data broker registry. I think there are, like, 600 different companies on it, and you can get the sort of things that are valuable for this sort of attack. You can find out family members.

You can go to dozens of data brokers and find out who family members are. You can get location. Sometimes you can get precise geolocation. You can find out about interests or what people have done. And so, you know, these little facts—these little factoids—are things they can say, oh, you were at the Braves game three weeks ago.

You know, you were the one millionth fan for the year and you won a big prize and we are going to wire you the money—give us your routing number.

So all of these sort of things can be used by hackers and I think that, as you all talked about now, they can do it at scale. AI can just kind of automate it all for you—automate the script, automate the e-mail, automate the—whatever the attack is. Like you say, who knows what they are going to be.

Strong legislation, privacy legislation, can help kind of rein that in. Like I said, about 20 states now have passed legislation, Colorado being one of the first. Those are great advances but it still puts a lot of onus on users. Users have to kind of go out of their way to track down these 600 data brokers and opt out. That is hard to do.

Even if we can opt in solutions are hard. You need permission and that is very difficult. So something that protects by default so you do not have to put all the onus on people to track down their data is something that we would recommend.

Senator HICKENLOOPER. Right. I agree completely. I think it is been frustrating to almost everyone. This is normally a large committee but frustrating that the data privacy continues to be elusive. It is hard to imagine.

Dr. Farid, you discussed different tools and techniques available for detecting AI-generated media including, you know, humans manually labeling content and automated detection tools.

How effective are the different kinds of tools and techniques to detect deepfakes and, you know, all these things from voices to video how well are they working?

Mr. FARID. Yes. So the first thing to understand is as a consumer your ability to distinguish reality from AI is becoming increasingly more difficult. I do this for a living and I am pretty good at it, and it is getting really hard.

So putting that onus on the consumer is look for the hands, look for this, look for that. That is a false sense of security that will simply not be true tomorrow, let alone three months from now.

In terms of automation of tools, there is two basic approaches to automation, what we call proactive and reactive.

Proactive, you have already heard of it from Mounir and Truepic. This is you are at the point of creation. Whether it is an AI-generated voice, an AI-generated image, or an image that I took with one of these devices in my hand, you can cryptographically sign the source of that so you know provenance.

That today is the only tool that works, and this is the important part, at scale. Your question needs a little bit of additional thing, which is at what scale do you want to operate, and if you want to operate at the scale of the Internet with billions of uploads a day, this today is the only technology that works at that scale.

The reactive techniques say, well, we have a bad actor who is not using credentials, or the credentials have not been deployed widely. So we wait. We wait to get the phone call.

We wait for the image to show up on social media and we respond to that. And there the answer of efficacy is it depends. If you give us enough time we will figure it out. We are pretty good at this.

But the half life of a social media post is measured in about 30 to 60 seconds. So it is not like you have a lot of time once it gets onto the social media platforms for you to respond.

So my view of these things is we need the proactive solutions. We need the reactive solutions to clean up the mess that is left behind and for consumers I think it is not so much about educating them on how to detect this but it is educating them on good digital hygiene on how to be careful and how to protect themselves.

Senator HICKENLOOPER. Great. I agree completely.

Ms. Mani, again, thank you for sharing your family's story. I know that it is not an easy thing and thank you for all your advocacy to address some of the harms that are being perpetrated through AI technology.

From your perspective, why is it so important to consider the perspectives of the people impacted by technology when we are looking at policy responses that address some of the harms to individuals?

Ms. MANI. I feel I have to deal with perspectives, many times misinformation. So education equals prevention. I feel it is—should be at forefront in any type of platform. So regulations in school, legislations in the government, as well as conversations with our students—children.

I think it is equally important to send a message to our girls that they are worthy and they should demand certain respect. But at the same time we need to hold the bad actors accountable and educate them as well.

We all know that the biggest protection is just the fear from going to jail. So, I mean, at this point FBI released this year statements saying that deepfakes are sexually explicit, deepfakes are illegal, and it is just surprising that our educational department did not follow with any regulations.

We all know what has happened right now in Pennsylvania to 46 girls and it is really not a unique situation. It has been happening and it will be happening unless we are going to start educating our youth because they are our future.

Senator HICKENLOOPER. Absolutely. I have to go vote so I am going to excuse myself and turn it over to Senator Luján from New Mexico. But I will be back. I have more questions.

STATEMENT OF HON. BEN RAY LUJÁN, U.S. SENATOR FROM NEW MEXICO

Senator LUJÁN [presiding]. Thank you, Mr. Chairman.

I want to recognize our Chair and our Ranking Member for this important hearing and to each of our panelists who are with us today.

According to the latest U.S. census data over 43.4 million people speak Spanish at home, making it the second most spoken language in America. This holds very true in my home state of New Mexico.

My question is for Dr. Farid. How accurate are current deepfake detection tools for audio and video deepfakes in Spanish compared to English?

Mr. FARID. Good. It is a good question. There are two different categories of detection schemes. So first of all, the proactive techniques are agnostic as to the language, right. They are there at the creation. They do not care what language you are speaking. So the kind of stuff that you will hear from Truepic does not matter.

The reactive techniques there are two categories. There are ones that are language dependent and there are ones that are language agnostic. We like the language agnostic ones. It is a big world and there is a lot of languages out there, and English is a relative minority. So many of the techniques that we have developed do not care about the language and the accuracy stays, roughly, the same.

The accuracies that you should expect, however, in the wild when you are dealing with things that are coming through your phone or what you are absorbing on social media on a good day 90, 95 percent accuracy. So you are going to misfire and that is true whether it is in Spanish or English.

There are ones that are language specific. They fired around the same accuracies, that 90 to 95 percent accuracy.

Senator LUJÁN. What is driving these differences in performance and what is needed to reach parity, understanding nonagnostic tools out there as well? Is there more that can be done not just here in the United States but as we engage with the global partners in this space as well?

Mr. FARID. Yes. The bias comes mostly from the training data. Most of the content online is in English and, certainly, what most of academics look at is English speaking because most of this research is happening here in the U.S. I think the next most common language that we do see is Spanish and so it is a close second, for obvious reasons.

But I would say the gap is not as bad as you think it is because most of the techniques that we and others have developed are language agnostic.

I am worried about all kinds of bias in the system. By the way, this is not on my top 10 list.

Senator LUJÁN. I appreciate that, sir.

Mr. Ibrahim, do the C2PA standards online how to display content origin information in Spanish so that Spanish-speaking consumers can understand whether the content they are viewing is AI generated?

Mr. IBRAHIM. Thank you, Senator.

So the mechanism of provenance, as Dr. Farid mentioned, is largely mathematical, right, so you are going to see the content credential pin on the digital content. I do not know if you saw some of the sample images we had earlier, but that aspect of it would, largely, be fairly standard across languages.

However, the information—the spec itself and information about our spec is something we want to move toward is internationalization.

We have done—we have spent a lot of time this year within the C2PA to raise awareness internationally because this is not just an American or English language standard. This is, in fact, a global standard.

So one of the things we did in October is we worked with the government of France at the French Embassy here in Washington to raise awareness for, you know, the G–20 community and other countries that are interested in looking at digital content provenance.

So internationalizing it not only in Spanish language but in parts of Asia, Africa, et cetera, is something we are certainly focused on.

Senator LUJÁN. It is my understanding that displaying of this information the position widely is that it should remain voluntary. Do you agree with this approach? Why or why not?

Mr. IBRAHIM. Yes, Senator, for authentic capture. So if I have my smart phone and I am taking an authentic picture or video absolutely it is my choice if I want to add content credentials.

Now, for Gen AI the platforms themselves have the provenance worked in. So all the outputs—think of it almost as an assembly line—are getting cryptographically hashed with this standard.

So I do believe it is a slightly different approach for gen AI or AI, but for authenticity and authentic capture 100 percent, sir.

Senator LUJÁN. One year ago at a hearing before the Senate Aging Committee, Gary Schildhorn testified about a call that he received that appeared to be from his son Brett pleading for help from jail. Mr. Schildhorn was so upset by the authentic seeming call that he just about sent the $9,000 that was being demanded from a fake public defender in bail money.

A 2024 FTC report found that older consumers have experienced an estimated $61 billion in financial loss due to scams over the last year. Older consumers lose more money to scams on average compared to younger consumers. Nearly one in 10 adults over the age of 65 have dementia. Approximately 25 percent of older adults live with mental health conditions.

Older consumers are a clear target of phone scams, personalized phishing e-mails, and other deepfake tactics.

Mr. Brookman, what is the most important thing that policymakers must do to protect older consumers from being scammed through these AI enabled tactics?

Mr. BROOKMAN. Yes, thank you for the question.

It is—yes, there have been lots of academic research showing that older Americans, even setting aside dementia, are more likely to fall prey to these schemes. I think anecdotally we probably all recognize that.

My own grandfather called me and said, hey, I got an e-mail from FrenchLottery@yahoo.com—I am a billionaire now. I said, no, papa, you are not.

But I think—like, I mean user education is, sadly, the most important thing, like, talk to your family members. The government can try—should be doing more. Senator Blackburn mentioned the FTC says it is a $10 billion problem. That probably massively understates that. I am sure it is much greater than that.

So we need to expand the resources to teach people about it and it is going to be a constantly changing battle because, again, these techniques are going to get more and more sophisticated.

So each of us individually should be talking to our family members but the government should be investing a lot more in education.

Senator LUJÁN. I appreciate that very much.

I recognize the Ranking Member of the Subcommittee for questions, Senator Blackburn.

Senator BLACKBURN. Thank you, and thank you all for the discussion. I know that Senator Hickenlooper asks about data privacy and the importance of getting that done so that people have the ability to protect their presence online.

And, Mr. Farid, I would like to just ask you, when you look at this issue what kind of added protection could having a Federal privacy standard provide before people get into so much of this AI-driven content?

Mr. FARID. Yes. I wish I had good news for you but I do not.

So first of all, there are about 250 million people in this country and our data is out there and there is no bringing that data back in.

So maybe we can protect the next generation, the one after that. But the 350 million of us that have been on the Internet for a few years that game is over.

Here is the other problem is if you look at gen AI today, to clone your voice I need about 20 seconds of audio. That is not hours of data that are being mopped up. It is 20 seconds, and soon it will be 10 seconds and soon after that it will be five, which means somebody can be sitting next to you in a cafe and record your voice and they have it.

To put you into a deepfake imagery—one image. You got a LinkedIn profile? Done, right? So how do you protect somebody when it is not about vacuuming up tons and tons of data?

Do not get me wrong. It is—should be criminal that we do not have a data privacy law in this country but I do not think——

Senator BLACKBURN. It would have given us those protections years ago.

Mr. FARID.—it would have helped.

Senator BLACKBURN. We have tried for 12 years and big tech has fought it.

Mr. FARID. Yes, they have.

Senator BLACKBURN. So it is truly frustrating.

Mr. Brookman, with Consumer Reports you all have quite a reputation with consumers coming to you for information. So talk just a little bit on how you educate people?

Mr. BROOKMAN. Yes. So we have a lot of tools. We have a tool called "Security Planner" that we try to franchise out to folks. We try to make a lot of our guidance available in Spanish language as well to make folks like this aware.

Like, what are the basics—what are you worried about? What is your threat model? Are you worried about an ex-boyfriend? Are you are you worried about a scammer?

And we tried to walk them through, like, you know, what they should do. You know, look for the warning signs. false urgency, right—like, you know, we need the money right now.

Anyone asking being paid through dodgy means. Like, if your boss is asking for, like, $500 of gift cards from CVS that is probably not actually a legitimate thing. We try to—you know, and so we try to give practical guidance.

Again, we cannot make it too overwhelming for folks like this cannot be a full-time job for folks. We try to give people the basic information that they have and they need to get through their day.

Basic cyber security practices, having different passwords because if you do get hacked on one site you can use a different pass phrase on another site.

One thing that could potentially help for some of these grandparents scam attacks is having a family safe word so only—it is only an accident if you say rutabaga or whatever your family's safe word happens to be.

Again, this does put a lot of onus on individuals but, unfortunately, that is the world that we live in and so we all have an obligation to be a little bit more—a little bit more savvy.

Senator BLACKBURN. Yes, and I know some of the senior specific organizations are beginning to do education because so many people are now using at home health and other conveniences, that much of their data is being transmitted and sensitive data—things that are HIPAA protected.

So, Mr. Farid, I want to come back to you on—as you are looking at this, and we say 5 years from now, 10 years from now, because we look at the fact that, as I just said, privacy, we have never gotten that Federal standard.

So people have treasure troves of information as out there. So when you look at advancements and changes in AI five years down the road what do you expect? Ten years down the road what do you expect?

Mr. FARID. Everything is going to get worse and, by the way, I do not think it is 5 years. Five years—think about—here is something to think about. ChatGPT went from zero to 1 billion users in one year.

Five years is an eternity in this space. We need to be thinking tomorrow and next year. But here is what we know. Hundreds and hundreds of billions of dollars are being poured into predictive AI and generative AI.

The technology is going to get better, it is going to get cheaper, and it is going to become more ubiquitous, and that means the bad guys are going to continue to weaponize it unless we figure out how to make that unbearable for them.

I think the threats are continuing, and that is not—I am not extrapolating, right. We have seen this from the very early days of this AI revolution. From the very first days we have seen the weaponization of this technology and it is only getting worse.

You cited the numbers in your opening statements. Those numbers will continue. The threats will continue.

Senator BLACKBURN. OK. Then you talk about making it a heavy price to bear so that means penalties and enforcement and involves law enforcement.

So, Mr. Ibrahim when we look at this as a global issue, what do you see as cooperation with some of our allies that are there?

I know many times when we are at the EU working on privacy, data security, things of that nature, they talk about our hesitancy to put these regulations in place.

So what kind of international cooperation do we need?

Mr. IBRAHIM. Thank you, Senator.

Indeed, I do believe that working across like-minded countries to establish at the very least best practices that we can all agree on, encourage private industry and technology and consumers to the extent they are following that to all align on.

We can educate together. We can work to establish at the very least a framework in which transparent information might be marked differently, elevated, because, as Dr. Farid and my colleagues here on the panel noted, we are not going to be able to stop every bad actor.

That is just not going to happen. You have open source models that are ungovernable, that are forked and are used by bad actors to do whatever that they please.

But what we can do is we can empower those that want to be transparent and that can change the dynamics of the digital content we see and hear online. We can—you know, social media companies, for example, can algorithmically design things so at the very least you are seeing transparent things without suppressing anything else.

So aligning with our partners, I think the G–7, the G-20, are great places to start on at the very least best practices, even if we do not agree on full regulations, for example, like the EU's AI Act. That might go too far for the American audience but there are areas of overlap.

Mr. BROOKMAN. Excellent. Thank you, and I yield back.

Senator HICKENLOOPER [presiding]. Thank you, and thank you for filling in, Senator Luján.

Senator Baldwin.

STATEMENT OF HON. TAMMY BALDWIN, U.S. SENATOR FROM WISCONSIN

Senator BALDWIN. I want to thank all of our witnesses for appearing here today to share your expertise and your personal experiences.

I believe it really greatly benefits this committee and the American public to hear how technology is being used to increase the sophistication and prevalence of crimes and scams.

I wanted to start with Ms. Mani. I especially want to thank you for sharing a deeply personal story with the public and with the Committee about your family's experiences with artificial intelligence deepfakes.

Would you be willing to share with the Committee the various steps that your family took after finding out your daughter's likeness had been used in this way? What sort of steps did you take and how did you know what to do and what to do next?

Ms. MANI. Well, thank you, Senator.

I am going to start with saying that our situation is really not unique. It has been happening and it is happening right now. I mean, we just heard about Pennsylvania, sorry.

So last year when we found out or when we were informed by the school what has happened to us the first thing we did, obviously, we called a lawyer.

In the school sector we were informed that nothing can be truly done because there are no school policies and no legislation, and the lawyers repeat exactly the same thing.

So when my daughter heard from the administration that, you know, in parenthesis she should be wearing a victim's badge and just go for counseling, and when she came home and she told me, I want to bring laws to my school so that way my sister—my younger sister—will have a safer digital future I said OK and that is how we started.

We have been advocating on multiple fronts just because as complex this technology is the context of misuse is very complex as well. So we have been looking out, obviously, at the legislation as our first road for change and surprisingly so it was the easiest one.

There are so many legislations right now that are—hopefully, will be passing soon and helping victims of deepfake misuse. We were looking in to the root of education as prevention, as setting an example of how victims should be treated.

The old tale of consent that I feel it needs to be reinforced even in 2025, but most importantly we start in acknowledging that it is time for us to change the rhetoric no matter where we were. Going and talking to whom we were talking was always about how do you feel as a victim, and even though it is very important how victims feel—it is an important part of the story—I feel it is time for us to start talking about the perpetrators and the lack of accountability and the need for stronger protections.

Like, for example—and as I mentioned, there are so many AI bills and so many AI laws out there and they are all important.

Some of them resonate deeper with us than others. Take it Down Act is very close to our heart for one important reason, which I feel public should be educated about.

Civil laws can be very expensive and not many can utilize them. Criminal laws—not everybody would like to go this route for multiple reasons—cultural, personal, religious, you name it.

The Take it Down Act allows the victims to take control over their own image and I think that is so important, which gives the freedom to anybody affected to just move on with their life, which sometimes that is all they want.

Senator BALDWIN. Thank you.

Mr. Brookman, in Wisconsin we are seeing a prevalence of technology being used to scam families including seniors. This past March the Dane County Sheriff's Office put out a warning about phone scams using artificial intelligent tools to mimic the voices and faces of loved ones to trick senior citizens into believing their loved ones are in danger and need help.

I know other law enforcement agencies in Wisconsin and across the country are hearing from victims of similar scams. I also hear directly from constituents on this matter.

In 2023, Wisconsinites lost approximately $92 million to fraudsters and scamsters. So once people realize they have become a victim of a financial fraud or scam can you speak to what steps people must take to report the crime and recover damages?

Mr. BROOKMAN. Yes. So, you know, first they are going to try to get the money back and it really depends on how they did it. If you paid someone with a credit card there is actually pretty good protections in this country.

If you used a peer to peer payment app, actually the protections are a lot weaker. I know there is some legislation before this committee to try to—because, again, it is a different button on your phone and suddenly the protections are a lot weaker.

If you paid with crypto or a gift card or a wire transfer it is going to be tough. And so, again, educating people just about those differences is really important. Freezing your credit can often be a good idea, again, depending on what got compromised.

It has gotten a lot easier in some ways since the Equifax data breach but still three different bureaus. If there are family members you need to get to that is still labor you need to do.

Different law enforcement—it is a good idea to report. Maybe report to local law enforcement, report to the FTC, or report to your state attorney general. If you get locked out of your accounts, like, that can be just an absolute battle to get back in if the scammer takes that over.

I think Matt Honan famously wrote a really detailed story about how he lost access to Gmail and the journey to get back there.

So it can be absolutely overwhelming and dispiriting and it puts a tremendous burden on people who have already been ripped off.

Senator HICKENLOOPER. Great, and as the Chair of the Subcommittee I do have the discretion to ask more questions even though I had to race off to vote. So, hopefully, you will indulge me. I will not go on too long.

I realize everybody is busy and really do appreciate you making the time.

Dr. Farid, we talked a lot—a lot of people talk about watermarks and other ways of designating some way of telling authentic from synthetic content, at least from what the origin is.

But some kinds of watermarks can be manipulated, removed, impersonated. What are example for methods to make these types of designations more secure and how resistant can they be?

Mr. FARID. Yes. So when we talk about content credentials we have to go beyond watermarks. There is three legs to the stool here, right.

So the first leg is what is called metadata. So this is where you just simply attach text to the file associated with the image, audio, or video that says, I created this at this date and time and it is AI or it is natural, and that text can be added or removed.

So it is not the most resilient but, you know, 10 percent of the people do not know what metadata is and, you know, you will catch them.

The watermark is the second technique where instead of adding it to the file you literally embed it into the underlying content. So what is nice about watermarks and metadata is you are tagging content and then putting that content into the wild where it can be identified from the piece of content and only the piece of content.

But that upside is also the downside because if the tag is there somebody, a sophisticated actor, can reach in and rip out the metadata and yank out that watermark and different watermarks have different resilience.

The third leg to the stool is to extract a digital signature from your content. So if, for example, OpenAI creates an AI-generated image, insert the watermark, insert the metadata, and then pull out a signature that I will store server side, and what that allows me to do is if somebody attacks my watermark and my metadata I can reattach the credential.

So those three things—each one in isolation does not get you where you want to but the three, now to your last question, are good.

Are they perfect? No, but they lop off a nice chunk of the problem, and then those reactive techniques that I talked about plus the public education, plus the regulatory, plus all those other things start to fill in the gaps.

So I am all for it but we should be very clear that this is not bulletproof, that this can and will be attacked, right.

But we raise the bar. Everything in cybersecurity is mitigation strategy. We do not eliminate threats. We simply mitigate them.

Senator HICKENLOOPER. Do not say that. Do not say that, however true it might be.

And we have talked a lot about the—this notion of how we better inform customers.

Mr. Ibrahim, Truepic is a—obviously, a leading voice along with other industry partners in developing content provenance technology to better inform consumers about the content they see—what we were just talking about.

But what are the incentives to encourage more industry adoption of labeling standards and investments in related research?

Mr. IBRAHIM. Thank you, Senator. So from our perch the incentives—the largest incentives has been the scale and the proliferation of generative and AI material.

We have seen entire industries that were digitizing because of the efficiency and the cost savings well before the advent of AI, unable to compete in today's economy if they do not have authenticity, transparency, or provenance in the digital content that they are operating on.

So we have seen that as the greatest incentive and that is why entire industries like insurance, business credentialing, warranties, supply chain monitoring, et cetera, et cetera, are adopting this technology.

However, on the consumer facing side, you know, think peer to peer commerce, home rental sites, online dating sites. The incentives although are there there have not been the financial incentives, at least from our perch, or consequences, for these platforms to better protect or at least give more transparency to their consumers and that has been lagging.

In the past year we have seen four of the largest social media platforms—LinkedIn, all the Meta sites, YouTube and TikTok, I believe, all begin implementing at various levels and at various commitments the C2PA open specification and that is great. But much more could be done in this regard.

Senator HICKENLOOPER. Right. Interesting. That is—obviously, that is something this committee has got to pay attention to. Ms. Mani, you have discussed the need for increased AI literacy in our schools, really, at all ages, whether elementary school or all the way through high school.

How do you calibrate what the top benefits are? I mean, as you look at it, when you look at benefits and risks how do you—what do you present to our students? What do you think they should learn?

What do they have to have top of mind? Let us put it that way.

Ms. MANI. Thank you, Senator. I think that is such an important question because I feel personally many of the misunderstanding comes from miseducation and misrepresentation and just lack of understanding.

So I think when we talk about AI, especially to our youth, we need to talk simultaneously about these pros and cons, right. It goes hand in hand.

So it can—it is an amazing tool that has been advancing our society right now for years in research, in education, in art, in medicine, et cetera, et cetera. At the same time, it needs to be regulated because it has repercussions if misused.

And I think it is equally important to educate about the repercussions to both the victims and the bad actors because at this point since it is illegal—and in many states we already have legislations—it can affect equally the victims professionally, personally, educationally if they apply for schools, emotionally and socially.

But it can also affect the perpetrators because they can end up in jail, and if we have no control over how certain children are raised at home or just simply, you know, they are too young to understand or comprehend, we must put things in a perspective so they understand the consequences.

And I also think, just to add to it, context is extremely important because we have K to 12 children who are right now misusing AI. We have college, which is a very different context, and we have perpetrators and pedophiles which should be looked from a singular perspective.

Senator HICKENLOOPER. Yes. No, I could not agree more, and consequences we need to——

Ms. MANI. Yes.

Senator HICKENLOOPER.—as we do that informing and proselytizing there is going to be consequences for that. Because you are right, people want to do the right thing but also people act properly because they have some consequence frequently.

Ms. MANI. Correct.

Senator HICKENLOOPER. Mr. Brookman, in your testimony you advocate for stronger enforcement, clearer rules of accountability for trying to mitigate customer harms and damage from these various scams and frauds.

I kind of love that word fraudster. Somehow it sounds softer. It is probably the wrong word. We should actually use a word that has a much harsher edge.

What actually do you think the FTC—the Federal Trade Commission—should take and what—if you were to have an opportunity what additional powers or authorities do you see as needed to protect the customers from all these various harms we have talked about today?

Mr. BROOKMAN. Absolutely. Thank you, Senator.

So I mean, again, I want to see more aggressive enforcement. Like, there were the five cases they brought in September. Those were very welcome.

And the cases against the tools when they are being used for bad like the—we talk about the scam—the fraudsters. The scammers are a dime a dozen. They are all over the world.

But if they are all using the same tool and the tool could really only logically be used for harm and that was one of the cases that the FTC brought, like, that can be very powerful.

I think we need to see more staffing. Again, they are considerably smaller than they were in 1980. They need more technologists. I was actually part of the first technology group that was created at the FTC, the Office of Technology Research and Investigation, but it was only, like, five people.

I think it has expanded somewhat under Chief Technologist Nguyen but, like, they still need more resources.

Two more things. There is, like, penalties. Like, when someone breaks the law they need to pay a fine. That seems, like, very basic. I worked for an attorney general. We had that authority. All attorney generals have it. FTC does not have it most of the time.

And then restitution. I mean, like, you know, since the AMG Supreme Court case, if they catch a fraudster who, like, stole, like, $20,000—$200,000—they often lack the legal authority to make them even just give that back. Not even penalties. Just give the money back.

I know there have been hearings, I think, in this committee and in the House on restoring the FTC's 13(b) authority. I think there had been bipartisan consensus around it.

But since—after the initial being upset that it was taken away I think that it has fallen behind as a priority, which is too bad. There is actually a news story that came out just yesterday that the FTC said we got a lot less money back for consumers this year because we cannot get money back for consumers.

So I think that needs to be fixed.

Senator HICKENLOOPER. So great. Thank you.

Senator Sullivan from Alaska.

STATEMENT OF HON. DAN SULLIVAN, U.S. SENATOR FROM ALASKA

Senator SULLIVAN. Thank you, Mr. Chairman, and I am going to just follow up on this line of questioning because I think it is a really important one.

I want to start with you, Ms. Mani. I am really sorry about what happened to your daughter. I have three daughters and, you know, it is horrendous. I can only imagine what you guys are going through.

So I want to start with you but I am going to kind of pose the question to everybody once you answer, and it just goes to what Senator Hickenlooper was asking about—I think it is what we are all struggling with—and it is responsibility and consequences, and when the responsibility and consequences are too diffuse then there is nobody who is deterred from taking this kind of action.

So in the 2022 Violence Against Women Act Congress passed legislation that created a civil cause of action for victims to sue individuals responsible for publishing nonconsensual intimate imagery.

So that is different from the deepfakes but it is kind of along the same lines. So my question as I am thinking through this, the consequences could fall on a number of different entities—the AI platform that generates a deepfake of, you know, someone's daughter and something that is fake but, you know, looks real; the social media platform that posts it and I would welcome to hear about your experience on trying to get them to take it down.

Because just having constituents in Alaska who deal with this, trying to get these companies to take something down that is harmful can be very challenging.

And then, third, of course, the person who was behind it. In this case I think it was a classmate of your daughter's, who only received a brief suspension. So all of these are different kind of potential parties to be responsible.

So I would ask for your view, first, in your specific situation and then I would ask all of you—I am a big believer in deterrence, right. You lay it out. Hey, if you do this you are going to get hammered. Hammered.

You are going to go to jail. You are going to spend 5 years behind bars thinking about it. That creates an incentive for people to not do these kind of things.

But right now I think the system is set up in a way that is so diffuse in terms of who is going to pay consequences that it enables everybody to just ignore them.

So if you may—if you would like to try to address a few of those in your own situation with what happened to your daughter, and then the rest of the witnesses however you want to address that question.

Ms. MANI. Thank you, Senator. And I think that is such a such a broad question and I think it depends on the context. Each incident has a completely different context and should be looked from its own perspective.

In our case, and I think that is why, you know, talking about taking it down, I think that is why we are so fiercely advocating for Take It Down Act because it allows the victims to take ownership.

Senator SULLIVAN. And did that happen in your case or did it take a long time?

Ms. MANI. So our situation was slightly different because we have never had the image. We have never seen the image.

But I can tell you that I am in close contact with Alison from Texas with her mom, and Ali—we testified together and our daughters did as well.

And in their case, I know they have been—they have the image, maybe a few images, actually, and they were contacting Snapchat and asking them to take it down for eight months, and until Senator Cruz reached out personally it has now been down. After that, it took them literally 24 hours to put it down. So I think accountability——

Senator SULLIVAN. But it should not take calling—I mean, Senator Cruz is a great friend of mine. He is soon going to Chair this committee. But it should not take a phone call from a United States Senator to get one of these companies to act. Trust me, I have made those phone calls.

Ms. MANI. It should not take a Senator or a Congressman and it should not take a law. It should be an ethical responsibility of every platform and just to put things——

Senator SULLIVAN. Yes. I think with some of these companies they need a law.

Ms. MANI. Some of them do.

Senator SULLIVAN. You cannot rely on their good kind of moral standing.

Ms. MANI. Yes. But I think that is a clear example of how easily this can be mitigated.

Senator SULLIVAN. Yes.

Ms. MANI. Because if there is a will there is a way, you know.

Senator SULLIVAN. Yes. Did your—did the perpetrator—and if you do not want to talk about it I understand because, you know, I am sure it is difficult on your family.

But, like, did the perpetrator in the instance with your daughter—was there any punishment for the student? I heard he was suspended for a day or something pretty minor.

Ms. MANI. That is correct, and I feel that is where lies a big responsibility, too, on our educational system——

Senator SULLIVAN. Yes.

Ms. MANI.—that they educate our boys and girls just in general, our community, on what will be acceptable and what is not. And yes, they are hiding behind the lack of laws. But every school has a code of conduct——

Senator SULLIVAN. Right.

Ms. MANI.—that they fall into and it is simply each administration's decision of how they are going to handle an incident.

Senator SULLIVAN. Great. Any other thoughts from the witnesses? I know I am already out of time but it is a general question. I think it is a really important one.

Who should we make pay the consequences? And I think that helps with regard to the whole issue.

Mr. BROOKMAN. I have thoughts.

So, I mean, I strongly agree with you. Deterrence is incredibly important. Everyone on the staff probably has some responsibilities, certainly the perpetrator. But the people who make these tools, like, some of these tools have very, very limited positive societal use cases. To create intimate images based on a real person—voice cloning.

Like, I mean, maybe some very edge cases but overwhelmingly likely to be used for harm. If you are putting a product out in the market that is overwhelmingly likely to be used for harm you should bear some responsibility. That was the FTC's Ryder case.

Senator SULLIVAN. What would you say—like, when you say there is very few of these products that have a societal good what would you—give me just a quick list of the ones you are thinking about that do not have a societal good?

Mr. BROOKMAN. Yes. I mean, I think, like, generating innovative images based on a real person. Like, I can imagine some very theoretical cases but 99 times out of 100 will be used for evil.

Voice cloning maybe, like, one in 30. But, like, I mean, again, maybe, like, someone—like a voice actor wants to franchise their voice. Maybe a person who is losing their voice wants to maintain it for sort of longevity.

But overwhelmingly they are likely to be used for scams. You know, at the very least, like, you know, get a—make them read a script to say this.

Image generation maybe just should not be available. Maybe it should just be per se harmful and illegal to offer that tool.

Finally, like, the social media sites this has been an issue, like, I have been working on for 20 years. Like, back when Facebook was a baby company I was in an attorney general's office.

We would send in complaints, like, hey, I am an underage user being sexually harassed—do something about it, and it was, like, crickets.

And then we brought a settlement against them to make them promise to do things within, like, 24 hours. Over time they have dedicated more resources to it. Have they dedicated enough resources to it?

I would say no, and so maybe clarifying they have some more obligation to respond to these things seems absolutely worthwhile.

Senator SULLIVAN. OK. Anyone else on the witness——

Mr. IBRAHIM. I would just add, Senator, one question that comes to my mind, particularly when it comes to distribution channels—the social media channels where a lot of this bad actions are taking place—are they doing everything they can from a technical perspective to try and mitigate that.

While maybe you are always going to have bad actors who use open source models and create deepfakes, but there is a rising problem of sextortion, particularly on social.

There are things that some of these social media companies can do to mitigate those chances. It will never eliminate it completely. Are they taking those actions and asking the question, why not. I think by itself would be critical.

Senator SULLIVAN. OK.

Mr. FARID. I will add a few things here. First of all, I believe in action and consequences and I think that is whether it is an individual, a trillion-dollar company or a U.S. Senator, for that matter, and I think the fact is that Silicon Valley and trillion-dollar companies have gotten away with it for 25 years and, frankly, we have let them.

Now, I think with the specifics I think it depends on the content. I think if it is child sexual abuse material, the perpetrator, the AI company, the tech company that is hosting it—everybody goes to jail.

This is easy, right? I think when it comes to nonconsensual imagery I think it is more complicated. I think if you are a 12- to 14-year-old and somebody has handed you a tool to create porn. Are we surprised that this is exactly what they do?

Do we think that 12-year-olds should go to jail for this? No. I think there should be consequences but I do not think it should be jail. But I think the person who created the tool and advertised it to say, take women's clothes off, and I think the financial institutions that tag on—a Visa and MasterCard tag that says pay for it with Visa and MasterCard—and I think the Googles of the world that surface that from a web search all have responsibility for putting that power into that 12-year-old.

And then, of course, with fraud it is the same thing. It is not just the person creating it. It is not just the tool. These things are being not just hosted on Facebook, on TikTok, on YouTube, on Instagram, on Twitter. They are being algorithmically amplified by those platforms because they monetize it.

They have a responsibility for that and we are letting them off the hook. So up and down the chain I think people have to be held accountable and until we change the calculus the fact it is not reining in abuses is good for business.

Senator SULLIVAN. Yes.

Mr. FARID. And we have to make it bad for business.

Senator SULLIVAN. Good. Great. Good answers. Thanks, panelists. Thank you, Mr. Chairman.

Senator HICKENLOOPER. Thank you, Senator.

Now we have on Zoom, Senator Klobuchar from Minnesota.

STATEMENT OF HON. AMY KLOBUCHAR, U.S. SENATOR FROM MINNESOTA

Senator KLOBUCHAR. All right.

Well, thank you so much, Chairman Hickenlooper, for having this hearing, as well as Senator Blackburn, and thank you to Chair Cantwell and as well as Ranking Member Cruz.

I care a lot about this subject. I guess I would start with you, Ms. Mani.

As you know, Senator Cruz and I lead the Take It Down Act and we have actually passed the bill through the Committee—through the Commerce Committee—and I am feeling very good. We had some issues we had to work out on the floor and I think we have resolved these. So we can actually pass this bill.

And I think you know the reason for this more than anyone. Just when you look out from your personal situation in 2016, one in 25 Americans reported being threatened with or being a victim of revenge porn.

Now eight years later that number is one in eight. Meanwhile, the proliferation of AI-generated deepfakes is making this problem worse. Approximately 96 percent of deepfake videos circulating online are nonconsensual porn.

So, I know—and you testified about your daughter and I am so sorry about what happened to her. And, as you know, our bill would ban the nonconsensual publication of intimate images—real or deepfake—and require the platforms to take the images down within 48 hours notice.

In your testimony you mentioned that schools have cited the absence of state and Federal laws as a reason for not taking action when kids are victimized. How would a Federal law in this have changed the experience—the horrific experience that your daughter and your family went through?

Ms. MANI. Well, thank you, Senator. I think I am going to just start with saying that in our case—in our high school it will allow the platform for the school to act and do anything at this point.

In schools in general I feel, as I mentioned it before, laws are form of education to our society, especially right now that we are dealing with schools and the problem of deepfakes over there.

They are a form of education our society of what is acceptable and what is not and what has not been delivered at home or in school can be delivered through laws and by fear of being criminally or in a civil way anyway affected by their actions, will allow, I guess, or will prevent, at least in some instances, from deepfakes being created.

At the same time, as I mentioned before, criminal laws and civil laws even though they are so important not always the victims will choose to use them in their advocacy for their image.

But the 48 hours take it down component of your bill is something that gives the victims immediate way to take the ownership of their image and some of them that is what they want.

Senator KLOBUCHAR. Very good. Thank you. I appreciate that.

Another difficult story is my own—one of my own employees her son is in the Marines and her husband got a call. Someone had scraped the son's voice off the Internet and left this very believable message and talked to the dad and said, "Dad, I need help. I need help. I need money."

And the dad thought it was suspicious because where he was stationed he was not allowed to call. And anyway those we have since—I have looked into this a lot and we have and it was, of course, a fake call, and we are starting to see families of people in the military preyed upon and it only takes a few seconds of audio to clone a voice using AI.

Criminals can pull the sample from public sources, as we know, like social media. As a result, AI-enabled scams are becoming far too common.

Dr. Farid, while there is technology to detect synthetic images and videos I am concerned we are behind on finding ways to detect a synthetic voice when it is heard over the phone.

How can the Federal Government best leverage available technology to verify the authenticity of audio, particularly in cases where a consumer does not have access to meta data?

Mr. FARID. Yes, Senator, you are right to be concerned.

The problem with the telephone is the quality of the audio that comes over is quite degraded as opposed to, for example, a YouTube video or a TikTok video and that inherently makes detection very difficult.

Also, there are serious privacy concerns. Are we going to listen to everybody's phone call and monitor that for a deepfake or not?

So I think the burden here has to shift to the producers. If you are an AI company—and you heard this from my colleague here—and you are allowing anybody to clone anybody's voice by simply clicking a box that says, I have permission to use their voice, you are the one who is on the hook for this.

So going after the telecoms. Look, we have not been able to get them to stop the spam calls so why do we think we are going to get them to stop the AI fraud? So I think we have to go after the source of the creation of these deepfake audios.

Senator KLOBUCHAR. All right. Well, thank you.

I see Senator Markey's face on the screen so I will forgo any additional questions. So thank you and thank you, Mr. Chairman.

Senator HICKENLOOPER. Thank you, Senator Klobuchar.

Senator Markey.

You got a hit. You are on mute. Senator Markey, you are on mute. You are still on mute. Cannot hear anything. Maybe you better talk to Senator Klobuchar.

Oh, there you are. You were there good for a second. Try again. You were there for a second.

STATEMENT OF HON. EDWARD MARKEY, U.S. SENATOR FROM MASSACHUSETTS

Senator MARKEY. Can you hear me now?

Senator HICKENLOOPER. Yes.

Senator MARKEY. Hello?

Senator HICKENLOOPER. Yes, we hear you. You are loud and clear now.

Senator MARKEY. Can you hear me now?

Senator HICKENLOOPER. Yes.

Senator MARKEY. Oh, beautiful. Thank you.

Our witnesses today have done a very effective job in demonstrating that artificial intelligence poses serious new risks for fraud and scams across different sectors. From AI voice cloning tools to deepfake images and videos AI is threatening to undermine our ability to tell truth from falsity, to sow distrust in our understanding of reality itself.

These scams and frauds will affect all Americans, but if history is any guide, marginalized communities are likely to be the greatest targets.

Mr. Brookman, in your written testimony you discuss a recent report by *Consumer Reports* that found black and Hispanic Americans were more impacted by digital attacks and scams.

Can you elaborate on your research?

Mr. BROOKMAN. Yes. This is, like, a report that we do every year to kind of track how people are adopting cybersecurity techniques, you know, using specific passwords.

This is the first year we asked about scams. You know, have you encountered scams—have you been—someone attempted a scam. Like, at least half of the people had, and of those who had, I think, overall about 10 percent of the people, the respondents to our nationally representative sample said they had lost money in a scam.

And the percentages I can get them for you but I think—I believe it was, like, twice as much, like, for Black and Hispanic people.

Of those who had encountered scams I think, like, a third of Black consumers and a third of Hispanic consumers had lost money and I think it was, like, about half of that for white Americans.

This is but—this is consistent with the research the Federal Trade Commission has done. The Federal Trade Commission has done a couple of reports looking at similar issues—also found that.

So I know we have talked a lot about in this hearing about the need to educate senior citizens and maybe many military families as well, and I strongly agree with that. But, like, we also probably need to reach out to marginalized communities too to—because that is where the money is.

That is where people are losing money—to make sure they are educated and prepared for this new wave of AI-powered scams.

Senator MARKEY. Thank you, sir. Thank you for your great work.

Because these issues now are well documented. In fact, in 2021 both AARP and the Federal Trade Commission published reports that determined that Black and Latino adults were more likely to be targeted by scammers and by fraudsters.

As artificial intelligence gives these bad actors new tools to target the public, communities of color will inevitably be the most impacted.

So, Mr. Brookman, do you agree that AI-enabled frauds and scams will have a disproportionate impact on marginalized communities?

Mr. BROOKMAN. Yes. I think the evidence so far indicates that. I mean, like, the traditional scams are more likely and so I think for AI-empowered scams as well it seems perfectly logical. So I mean, again, like, yes.

Senator MARKEY. Yes. So I agree because any effort to address AI-enabled fraud and scams must be spent given special attention to the unique harms that are facing those populations, and as we discuss AI's impact on marginalized communities we must also remember that AI-powered algorithms can supercharge preexisting bias and discrimination.

For example, a 2019 report found that due to bias in mortgage approval algorithms lenders were 80 percent more likely to reject Black applicants than similar White applicants.

On another occasion, a major tech company found that its AI résumé screening tools penalized résumés that included the words "women" and recommended male applicants for jobs at much higher rates than similar female applicants.

That is unacceptable. We cannot let AI stand for accelerating inequality and it is why in September, I introduced the AI Civil Rights Act which would ensure that companies review and eliminate bias in their algorithms and put Americans back in control of key decisions in their lives.

Mr. Brookman, do you agree that Congress should pass my AI Civil Rights Act?

Mr. BROOKMAN. Yes. I definitely agree that algorithmic harms from bias in existing inequalities is something that needs to be addressed. Dr. Farid testified about that earlier.

We have also supported legislation in—a law in Colorado, the first law in the country to address that. But yes, the Federal Government should do it and we have endorsed your bill on this issue specifically.

Senator MARKEY. And I appreciate that. We should not just leave it to the individuals to figure out how to protect themselves. We have to take action against the companies that set up these algorithms.

So as Congress considers AI legislation we just cannot ignore how those algorithms will impact marginalized communities and that is why we have to pass my AI Civil Rights Act and all of this, you know, requires that consumers be able to authenticate the digital content and we need to put those protections in place.

So I thank all of you for everything that you are doing to help to inform the public about these very important issues.

Senator HICKENLOOPER. Thank you, Senator Markey.

And before closing—I think we are out of Senators for now—I would like to ask unanimous consent that statements from Microsoft on the NCII and from the Center for AI and Digital Policy on consumer protection be entered into the record without objection.

I see no objection. So done.

[The information referred to follows:]

November 18, 2024

Hon. MIKE JOHNSON,
Speaker of the House,
U.S. House of Representatives,
Washington, DC.

Hon. HAKEEM JEFFRIES,
Democratic Leader,
U.S. House of Representatives,
Washington, DC.

Hon. CHARLES SCHUMER,
Senate Majority Leader,
U.S. Senate,
Washington, DC.

Hon. MITCH MCCONNELL,
Senate Republican Leader,
U.S. Senate,
Washington, DC.

Dear Speaker Johnson, Majority Leader Schumer, Minority Leader Jeffries, and Republican Leader McConnell,

This Congress has undertaken important efforts to advance legislation to address real and AI-generated non-consensual intimate imagery (NCII). Thanks to these bipartisan efforts, we are closer than ever to achieving what so many victims of this exploitation have long sought—true and effective remedies. That is why we are coming together to urge you to pass the Stopping Harmful Image Exploitation and Limiting Distribution (SHIELD) Act and the Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks (TAKE IT DOWN) Act this Congress.

These next few weeks present a critical opportunity for bipartisan unity, and this area has been one where Congress has already demonstrated that it can put aside political differences to help protect vulnerable people. In July, the Senate passed the SHIELD Act by unanimous consent. This bill establishes Federal criminal liability for individuals who share private, sexually explicit or nude images without consent. Likewise, the Senate Commerce Committee unanimously passed the TAKE IT DOWN Act this summer. That bill would criminalize the publication of non-consensual, sexually exploitative images—including AI-generated content—and require platforms to have in place notice and takedown processes. Both bills have bipartisan companions in the House.

We recognize that there is other legislation in this space that Congress may be considering. We are advocating that Congress pass these bills for three primary reasons. First, we are concerned about both real and synthetic NCII. While it is critical to address AI-generated and other synthetic content, Congress must also address the non-consensual dissemination of authentic imagery, which these bills do. Second, both civil and criminal relief are necessary pieces in legislation to deter bad actors, especially for victims who may not have the financial resources to secure justice on their own through a civil lawsuit. Lastly, combating this horrific content will take a whole of society approach and that includes platforms, who must take action to remove this exploitative, non-consensual content once notified. No other set of bills accomplishes these three goals like SHIELD and TAKE IT DOWN do.

Recent studies tell us the harm is dire, consequences are long-lasting, and the problem is not going away. In Microsoft's latest annual Global Online Safety Survey, respondents were asked about this topic, with 12 percent globally reporting having been exposed to sexual deepfakes in the last year. Fear about it is also increasing, with 73 percent of respondents globally citing concern about sexual or online abuse from AI. And yet, there also appears to be an alarming gap in understanding the exploitation involved. In the United States, only 56 percent of respondents thought it would be very harmful to use AI to create a false image of a politician, and only 63 percent of respondents thought it would be very harmful to create a nude image of a celebrity. While these bills might not solve all these concerns, they are a step in the right direction of taking action, protecting victims and holding bad actors accountable.

Together, we are proud to support the SHIELD and TAKE IT DOWN Acts. While we represent three different entities—one corporation, two non-governmental organizations—we are united in our desire to address this harm and protect people, especially women and girls, from online exploitation.

As Microsoft, we have long recognized our responsibility to keep our users safe, particularly children. We have a comprehensive policy addressing NCII harms, which has been in place since 2015. Microsoft does not allow the sharing or creation of sexually intimate images of someone without their permission. This includes photorealistic NCII content that was created or altered using technology. We do not allow NCII to be distributed on our services, nor do we allow any content that praises, supports, or requests NCII. In 2009, we collaborated with Dartmouth College to develop PhotoDNA, a hash-matching technology that was a landmark step forward in our collective ability to detect and address child sexual abuse material (CSAM) across the online ecosystem. In March 2024, Microsoft donated a new PhotoDNA capability to support the efforts of StopNCII.org, a service that enables people to protect themselves from having their intimate images shared online without their consent. In September 2024, we were proud to announce we have been piloting use of the StopNCII.org database to prevent NCII content from being returned in image search results in Bing. We will continue to evolve our approach, including as part of our comprehensive approach to addressing abusive AI-generated content. That approach is premised on a need for whole-of-society action, including steps to modernize legislation. To help better protect women and children from online exploitation, Microsoft was proud to endorse the SHIELD Act in our white paper, "Protecting the public from abusive AI-generated content", and to endorse the TAKE IT DOWN Act in July.

For over 40 years, The National Center for Missing and Exploited Children (NCMEC) has served as the Congressionally-designated resource center and clearinghouse on missing and exploited children issues. NCMEC has witnessed how offenders often misuse new technology to sexually exploit children online. Last year, NCMEC's CyberTipline received more than 36 million reports relating to child sexual exploitation. CyberTipline reports relating to the online enticement and sextortion of children have increased over 300 percent over the past 3 years. Recently, NCMEC has seen the emergence of new threats with offenders using generative artificial intelligence (GAI) technology to create nude and sexually exploitative images of children. Over the past 2 years, NCMEC has received more than 10,000 reports of child sexual exploitation that were created with GAI technology. It is essential that our laws continue to be updated to address these emerging threats to child safety online. The SHIELD Act will close a gap in current law by criminalizing the distribution of nude and exploitative images of a child. Similarly, the TAKE IT DOWN Act will criminalize the distribution of nude and exploitative images of a child that are produced with GAI technology. Both bills will close legal gaps and are essential to provide crucial legal remedies to combat offenders who are exploiting children online, especially with emerging new technologies.

As the Nation's oldest and most prominent nonprofit organization dedicated to combating image-based sexual abuse, the Cyber Civil Rights Initiative (CCRI) has been calling for a Federal criminal law prohibiting the nonconsensual distribution of intimate imagery for over a decade. CCRI provides the only 24-hour crisis helpline for victims and survivors of image-based abuse in the country, as well as an online safety center and other resources. Our advocacy efforts helped convince major tech companies to ban and remove image-based sexual abuse in 2015 and we continue to work with the industry to improve policies and practices relating to privacy and abuse. CCRI's cutting-edge research into the prevalence and impact of image-based sexual abuse has revealed that criminal penalties are the most effective deterrent against this abuse. Since 2013, our model legislation on NCII has served as the template for the wave of state law legislative reform that brought the number of jurisdictions criminalizing this abuse from 3 to 49 (as well as the District of Columbia, Puerto Rico, and Guam). CCRI's model legislation has also served as the template for Federal laws, including an amendment to the Uniform Code of Military Justice; a Federal civil provision passed as part of the Violence Against Women Reauthorization Act of 2022; and for SHIELD. While we celebrate the great strides that have been made with regard to state criminal laws and civil remedies for NCII, variations across jurisdictions in the definition, classification, and remedies for this crime have resulted in a patchwork of laws that is confusing for victims and law enforcement alike. Many state NCII laws are limited by misguided motive requirements, allowing perpetrators who commit this abuse for profit, voyeurism, or social validation to act with impunity. The SHIELD Act closes this loophole and serves as a powerful deterrent against this destructive and often borderless crime. The TAKE IT DOWN Act takes important steps to address the growing epidemic of sexually explicit digital forgeries and to create incentives for the removal of all forms of NCII.

Microsoft, NCMEC, and CCRI wish to express our gratitude to Senators Klobuchar, Cruz and Cornyn for their leadership on these critical bills, as well as to the other sponsors. We also thank you for your consideration, and we look forward to working with you to advance them this Congress.

Sincerely,

FREDERICK S. HUMPHRIES, JR.,
Corporate Vice President, U.S. Government Affairs,
Microsoft.

MICHELLE DELAUNE,
President & CEO,
National Center for Missing & Exploited Children (NCMEC).

MARY ANNE FRANKS, JD, DPHIL,
President and Legislative & Tech Policy Director,
CCRI.

CENTER FOR AI AND DIGITAL POLICY
November 19, 2024

Chairman JOHN HICKENLOOPER,
Ranking Member MARSHA BLACKBURN,
Senate Commerce, Science, and Transportation,
Subcommittee on Consumer Protection, Product Safety, and Data Security,
Washington, DC.

Re: CAIDP Statement for the Record: *Hearing to examine protecting consumers from artificial intelligence enabled fraud and scams*

Dear Chairman Hickenlooper, Ranking Member Blackburn, and Members of the Committee,

The Center for AI and Digital Policy (CAIDP) welcomes the hearing convened by this subcommittee on the urgent need to protect consumers from artificial intelligence (AI) enabled frauds and scams [1] Consumers today confront a wide range of new challenges because of techniques that mislead, manipulate, and misinform. This is bad for consumers, legitimate businesses, and the marketplace. The Federal Trade Commission (FTC) has a central responsibility to protect consumers against such unfair and deceptive trade practices. Although the FTC has issued useful business guidance,[2] the consumer agency has failed to enforce the necessary AI guardrails to safeguard consumers.[3] We submit this statement and request you to:

1. Urge the Federal Trade Commission (FTC) to conclude the investigation of OpenAI and issue an Order, following the detailed complaint CAIDP filed with the consumer protection agency in March 2023
2. Move forward bipartisan legislation that would safeguard consumers, including S.3312-Artificial Intelligence Research, Innovation, and Accountability Act (ARIA) 2024,[4] S. 4769—the Validation and Evaluation for the Trustworthy Artificial Intelligence Act,[5] S. 4178-Future of Artificial Intelligence Innovation Act of 2024,[6] and H.R. 5077—CREATE AI Act[7]

About CAIDP

The CAIDP is a non-profit, independent research and education organization based in Washington D.C.[8] In March of last year, we sent filed a detailed complaint with the FTC urging the agency to establish necessary safeguards for American consumers.[9] In July 2023, both the New York Times and the Wall Street Journal re-

[1] Senate Commerce Committee, Subcommittee on Consumer Protection, Product Safety, and Data Security, *Hearings to examine protecting consumers from artificial intelligence enabled fraud and scams,* 118th Congress (2023–2024), (Nov. 19, 2024), *https://www.congress.gov/event/118th-congress/senate-event/336302?s=1&r=1*

[2] FTC, *Using Artificial Intelligence and Algorithms,* Business Guidance, (Apr. 8, 2020), *https://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-and-algorithms;* FTC, *Aiming for truth, fairness, and equity in your company's use of AI,* Business Guidance, (Apr. 2021), *https://www.ftc.gov/business-guidance/blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai;* FTC, *Keep your AI claims in check,* (Feb. 27, 2023), *https://www.ftc.gov/business-guidance/blog/2023/02/keep-your-ai-claims-check;* FTC, *Chatbots, deepfakes, and voice clones: AI deception for sale,* Business Guidance, (Mar. 20, 2023), *https://www.ftc.gov/business-guidance/blog/2023/03/chatbots-deepfakes-voice-clones-ai-deception-sale;* FTC, *The Luring Test: AI and the engineering of consumer trust,* Business Guidance, (May 1, 2023), *https://www.ftc.gov/consumer-alerts/2023/05/luring-test-ai-and-engineering-consumer-trust;* FTC, *Consumers Are Voicing Concerns About AI,* Business Guidance, (Oct. 3, 2023), *https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/10/consumers-are-voicing-concerns-about-ai;* FTC, *Succor borne every minute,* Business Guidance, (Jun. 11, 2024), *https://www.ftc.gov/business-guidance/blog/2024/06/succor-borne-every-minute;*

[3] CAIDP, *ChatGPT and the Federal Trade Commission: Still No Guardrails,* (Jul. 2024), *https://www.caidp.org/app/download/8520338863/CAIDP-Still-No-Guardrails-July2024.pdf*

[4] Text—S.3312—118th Congress (2023–2024): Artificial Intelligence Research, Innovation, and Accountability Act of 2023, S.3312, 118th Cong. (2024), *https://www.congress.gov/bill/118th-congress/senate-bill/3312/text.*

[5] Text—S.4769—118th Congress (2023–2024): VET Artificial Intelligence Act, S.4769, 118th Cong. (2024), *https://www.congress.gov/bill/118th-congress/senate-bill/4769/text/is.*

[6] Text—S.4178—118th Congress (2023–2024): Future of Artificial Intelligence Innovation Act of 2024, S.4178, 118th Cong. (2024), *https://www.congress.gov/bill/118th-congress/senate-bill/4178/text.*

[7] H.R. 5077—118th Congress (2023–2024): CREATE AI Act of 2023, H.R. 5077, 118th Cong. (2024),*https://www.congress.gov/bill/118th-congress/house-bill/5077.*

[8] CAIDP, *https://www.caidp.org/about-2/*

[9] CAIDP, *Complaint to the FTC—In re OpenAI and ChatGPT,* (Mar. 30, 2024), *https://www.caidp.org/app/download/8450269463/CAIDP-FTC-Complaint-OpenAI-GPT–033023.pdf;* CAIDP, *Supplement to the Original Complaint to the FTC—In re OpenAI and ChatGPT,* (Jul.10,

ported that the FTC had launched the investigation we requested.[10] But a year later, there is still no conclusion.

In September 2023, we also wrote to this Subcommittee to urge the establishment of transparency mechanisms for AI systems.[11] Earlier this year, we also wrote to the Senate Commerce Committee on protecting American privacy[12] and to pass Federal privacy legislation.[13] We urged the Senate Commerce Committee to pass the American Privacy Rights Act.[14]

We write now on the urgent issue of establishing guardrails for consumers through congressional action and agency enforcement under existing laws.

1. Urge the Federal Tarde Commission to conclude the investigation of OpenAI and issue an Order

Over a year ago, CAIDP filed a detailed, formal complaint[15] with the FTC about OpenAI, alleging that OpenAI had violated U.S. consumer protection law by releasing consumer products without sufficient safeguards. We described real risks to the public—giving bad investment advice or telling children how to lie to their parents.

We cautioned that commercialized generative AI services, such as DALL–E, GPT–4o, and OpenAI Five, among other publicly available products, will turbocharge risks to public safety, online child safety, cybersecurity, and consumer deception.[16] Those risks have already materialized. The release of ChatGPT–4o's voice assistant, "Sky," sparked public outrage due to its eerie similarity with Scarlett Johansson's voice.[17] Additionally, OpenAI's "GPTs," a custom chatbot store powered by OpenAI's models, has become filled with spam.[18]

Alarmingly, it was reported in May 2024 that OpenAI is considering extending the uses of its model to enable users to create AI-generated pornography and other explicit content.[19] *OpenAI on the one hand touts its mission statement of "safe and beneficial AI" and on the other hand in its own model specifications is extending use-cases of its Dall-E image generator to "responsibly* create erotica, extreme gore, slurs,

2023), *https://www.caidp.org/app/download/8466615863/CAIDP-FTC-Supplement-OpenAI-07102023.pdf;* CAIDP, Second Supplement to the Original Complaint to the FTC—In re OpenAI and ChatGPT, (Nov. 14, 2023),*https://www.caidp.org/app/download/8485816363/CAIDP-Supplement-FTC-OpenAI-11142023.pdf*

[10] New York Times, *F.T.C. Opens Investigation Into ChatGPT Maker Over Technology's Potential Harms,* (Jul. 13, 2023), *https://www.nytimes.com/2023/07/13/technology/chatgpt-investigation-ftc-openai.html;* The Wall Street Journal, *ChatGPT Comes Under Investigation by the Federal Trade Commission,* (Jul.13, 2023), *https://www.wsj.com/articles/chatgpt-under-investigation-by-ftc-21e4b3ef*

[11] CAIDP Statement for the record: *The Need for Transparency in Artificial Intelligence,* (Sept. 12, 2023), *https://www.linkedin.com/posts/center-for-ai-and-digital-policy_caidp-statement-for-senate-commerce-committee-sept-activity*

[12] CAIDP Statement for the record: *The Need to Protect American's Privacy and the AI Accelerant,* (Jul. 10, 2024), *https://www.linkedin.com/posts/center-for-ai-and-digital-policy _caidp-senate-commerce-ai-and-privacy-july-*

[13] CAIDP Statement for the Record: *Mark-up hearing on H.R. 8188,* (Jun. 27, 2024), *https://www.linkedin.com/posts/center-for-ai-and-digital-policy_caidp-to-us-congress-on-apra-and-ai-june-activity_7211990172387680256-_5JI?*

[14] 118th Congress, 2nd Session, H.R. 8818, *American Privacy Rights Act of 2024, https://d1dth6e84htgma.cloudfront.net/ H_R_8818_American_Privacy_Rights_Act_of_***024***_a265f50b54.pdf; See also,* H.R. 8818—American Privacy Rights Act of 2024, *https://www.congress.gov/bill/118th-congress/house-bill/8818/text*

[15] CAIDP, *Complaint to the FTC—In re OpenAI and ChatGPT,* (Mar. 30, 2024), *https://www.caidp.org/app/download/8450269463/CAIDP-FTC-Complaint-OpenAI-GPT-033023.pdf*

[16] CAIDP, *Complaint to FTC—in re OpenAI and ChatGPT,* (Mar. 30, 2023), *https://www.caidp.org/cases/openai/;* Open AI, *The GPT–4 System Card* (Mar. 15, 2023), *https://cdn.openai.com/papers/gpt-4-system-card.pdf;* Merve Hickok, Marc Rotenberg, Christabel Randolph, *It's time for the FTC to act on ChatGPT,* Op-Ed, The Hill, (Jun. 14, 2023), *https://thehill.com/opinion/technology/4722343-its-time-for-the-ftc-to-act-on-chatgpt/* The Lancet, *ChatGPT: Friend or Foe,* Editorial (Feb. 6, 2023), *https://www.thelancet.com/action/showPdf*

[17] New York Times, *Scarlett Johansson's Statement About Her Interactions With Sam Altman,* (May 20, 2024), *https://www.nytimes.com/2024/05/20/technology/scarlett-johansson-openai-statement.html*

[18] TechCrunch, *OpenAI's chatbot store is filling up with spam,* (Mar. 20, 2024), *https://techcrunch.com/2024/03/20/openais-chatbot-store-is-filling-up-with-spam/*

[19] The Guardian, *OpenAI considers allowing users to create AI-generated pornography,* (May 9, 2024), *https://www.theguardian.com/technology/article/2024/may/09/openai-considers-allowing-users-to-create-ai-generated-pornography*

and unsolicited profanity"[20]—uses which have already affected many Americans through scams, and sexual exploitations.[21]

In September 2023, about 70 national and state consumer organizations called upon the FTC and Consumer Financial Protection Bureau (CFPB) to "protect consumers from the increasing threat of AI-generated "deepfake" voice clips and videos used for financial fraud."[22] The FTC has taken some steps, in line with its "Operation AI Comply" strategy to crackdown on AI-induced fraud and scam.[23]

As public scrutiny on OpenAI's business practices increase, one critical aspect is missing: action from the FTC. Overall, there have been *over a dozen* investigations of OpenAI in different countries, targeting various aspects of its services.[24] In our recent report ***"Still No Guardrails,"*** we called attention to the FTC's delay in concluding the investigation and establishing critical guardrails.[25]

AI does not operate in a vacuum. There are existing laws, and Federal agencies, including the FTC, have issued a joint statement stating their existing authorities over AI products and systems.[26] There is an urgent need to enforce consumer protection laws and enforce guardrails for AI systems, particularly those that directly interact with consumers or influence their behavior.[27]

We ask that this Committee urge the FTC to enforce consumer guardrails in the AI industry, to conclude the investigation of OpenAI and issue an order. While Congressional action progresses on broader AI policy issues, it is incumbent upon the FTC to enforce existing consumer protection laws to protect the American public from fraud, scams, and exploitation.

2. Move forward AI Bills—S. 3312, S. 4769, S. 4178, and H.R. 5077 on a priority basis during the lame-duck session

We appreciate Chair Hickenlooper and Ranking Member Blackburn in advancing several important pieces of AI legislation. Congressional action on AI has been a model for bipartisan cooperation.[28]

The Senate and House Committees have marked up several bills providing funding and resources[29] for the industry but with little or no accountability provisions. We need real accountability provisions that would strengthen guardrails for consumers in the context of AI systems.[30]

We support the passage of the Thune-Klobuchar S.3312-Artificial Intelligence Research, Innovation, and Accountability Act (ARIA) 2024[31] *(co-sponsored by Chair Hickenlooper and others) which requires transparency reports, risk management assessments, and enforcement against non-compliant high-risk applications.* As Sen.

[20] OpenAI, *Model Spec,* (May 08, 2024) *https://cdn.openai.com/spec/model-spec-2024-05-08.html*

[21] The Washington Post, *AI-Generated images of child sexual abuse are on the rise,* (Jun. 19, 2023), *https://www.washingtonpost.com/technology/2023/06/19/artificial-intelligence-child-sex-abuse-images/*

[22] U.S. PIRG and others, *Letter to FTC and CFPB,* (Sept. 13, 2023), *https://www.commonsensemedia.org/sites/default/files/featured-content/files/deepfake-based-financial-fraud-letter-to-cfpb-and-ftc.pdf*

[23] Alvaro Puig, *Operation AI Comply: Detecting AI-infused frauds and deceptions,* Federal Trade Commission (Sept. 25, 2024), *https://consumer.ftc.gov/consumer-alerts/2024/09/operation-ai-comply-detecting-ai-infused-frauds-and-deceptions*

[24] Stephanie Psaila, *Governments vs ChatGPT: Investigations around the world,* DIPLO (Jun. 16, 2023), *https://www.diplomacy.edu/blog/governments-chatgpt-investigations/*.

[25] CAIDP, *ChatGPT and the Federal Trade Commission: Still No Guardrails,* (Jul. 2024), *https://www.caidp.org/app/download/8520338863/CAIDP-Still-No-Guardrails-July2024.pdf*

[26] Joint Statement on Enforcement Efforts Against Discrimination and Bias in Automated Systems, *https://www.ftc.gov/system/files/ftc_gov/pdf/EEOC-CRT-FTC-CFPB-AI-Joint-Statement%28final%29.pdf*

[27] *See, e.g.,* CAIDP, Comment on *Trade Regulation Rule on Impersonation of Government and Businesses,* FTC Matter No. R207000 (Apr. 30, 2024); CAIDP, Comment on *Artificial Intelligence (AI) Systems Accountability Measures and Policies,* NTIA Docket No. 230407–0093 (Jun. 12, 2023).

[28] Marc Rotenberg, *A Turning Point for U.S. AI Policy: Senate Explores Solutions,* Blog@CACM, (May 17, 2023), *https://cacm.acm.org/blogcacm/a-turning-point-for-u-s-ai-policy-senate-explores-solutions/*

[29] Rebecca Heilweil, *Senate Commerce Committee advances several bills on AI,* Fedscoop, (Jul. 31, 2024), *https://fedscoop.com/senate-commerce-committee-advances-several-bills-on-ai/*

[30] CAIDP Statement for the record: *Oversight of AI: Insiders' Perspectives,* (Sept. 17, 2024), *https://www.linkedin.com/posts/center-for-ai-and-digital-policy_caidp-senate-judiciary-committee-ai-sept-activity*

[31] Text—S.3312—118th Congress (2023–2024): Artificial Intelligence Research, Innovation, and Accountability Act of 2023, S.3312, 118th Cong. (2024), *https://www.congress.gov/bill/118th-congress/senate-bill/3312/text.*

Thune has stated S.3312 "identifies some basic rules of the road that protect consumers."[32]

We also support the Hickenlooper-Capito S. 4769—the Validation and Evaluation for the Trustworthy Artificial Intelligence Act,[33] and the Cantwell, Young, Hickenlooper, Blackburn S. 4178-Future of Artificial Intelligence Innovation Act of 2024[34] and the bicameral Heinrich, Young, Eshoo, Obernolte H.R. 5077—the CREATE AI Act of 2023.[35]

Together these bills establish mechanisms for transparency and accountability for AI systems through funding and resources for safe innovation, technical evaluations, audits, assurances, and oversight. All of these bills have been reported favorably by the Senate and House Commerce Committees and it is imperative to move this forward.

Bi-partisan leadership is crucial, now more than ever, to bring these bills to the Senate floor for a final debate and vote on the provisions. ***We urge you-Chair Hickenlooper and Ranking Member Blackburn to work with Senate leader Schumer and McConnell to pass AI legislation, specifically, the aforesaid bills during the lame-duck session.***

Thank you for your consideration of our views. We ask that this statement be included in the hearing record. We would be pleased to provide you and your staff with additional information.

Sincerely yours,

MERVE HICKOK,
CAIDP President.
MARC ROTENBERG,
CAIDP Executive Director.
CHRISTABEL RANDOLPH,
Associate Director.
JANHVI PATEL,
Law Fellow.

Senator HICKENLOOPER. Let me just say thank you. I appreciate so much—I know you are every bit as busy as we are and that you took time out of your lives and schedule to come and share your experiences and your perspectives with us.

I think these are important discussions and it can be very frustrating to be in a body that moves at such a careful, slow pace when things that you are describing are so blatantly problematic and screaming for consequences, I think several Senators said.

I do think—just in the same sense that every great journey starts with the first step I think great legislation starts with the first hearing or maybe this might be the second or third hearing for some of you.

But I remain very optimistic and I do feel a great sense of urgency for all the reasons that you have each made clear.

So Senators will have until Tuesday, December 3 to submit questions for the record. Witnesses will have until Tuesday, December 17—our traditional two weeks—to respond to written questions.

And with that, this hearing is adjourned.

[Whereupon, at 4:23 p.m., the hearing was adjourned.]

[32] Sen. John Thune, *Thune, Klobuchar Lead Commerce Committee Colleagues in Introducing Bipartisan AI Bill to Boost Innovation and Strengthen Accountability,* Press Release, (Nov. 15, 2023), *https://www.thune.senate.gov/public/index.cfm/press-releases?ID=E27EFA8B-3AD1-42B5-8248-3987D2AFA649*

[33] Text—S.4769—118th Congress (2023–2024): VET Artificial Intelligence Act, S.4769, 118th Cong. (2024), *https://www.congress.gov/bill/118th-congress/senate-bill/4769/text/is.*

[34] Text—S.4178—118th Congress (2023–2024): Future of Artificial Intelligence Innovation Act of 2024, S.4178, 118th Cong. (2024), *https://www.congress.gov/bill/118th-congress/senate-bill/4178/text.*

[35] Actions—H.R. 5077—118th Congress (2023–2024): CREATE AI Act of 2023, H.R. 5077, 118th Cong. (2024), *https://www.congress.gov/bill/118th-congress/house-bill/5077/all-actions.*

APPENDIX

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. MARIA CANTWELL TO HANY FARID

Detection Technology

Last month in my home state of Washington, the Grays Harbor Public Utilities District warned that scammers are targeting consumers with voice cloning technology. According to Grays Harbor PUD, scammers are using this technology to create authentic sounding messages threatening to cut off utilities services and attempting to fool customers into sending money to avoid losing power.

Question 1. What types of technologies are being developed specifically with respect to deepfake audio detection to allow businesses and consumers to identify deepfakes, including scam calls?

Answer. Most of the efforts to detect deepfake voices is happening in Academe. There are, however, a few companies including a company that I co-founded (GetReal Labs) that is working to deploy these technologies to protect consumers and organizations.

AI Innovation to Fight Fraud

We know artificial intelligence is supercharging fraud. It is also one of our best weapons against fraud.

Question 1. How is artificial intelligence being leveraged to detect and prevent fraud?

Answer. The frustrating truth is that the largest tech companies are—for the most part—not particularly incentivized to develop and deploy these technologies. Put simply, safety and defense are not profitable. On the other hand, financial institutions are more incentivized, and we are starting to see a rollout of AI-powered technologies to protect consumers.

Question 2. How can the government support innovation in the fraud-fighting space and ensure that these innovations trickle down to consumers?

Answer. Hold AI companies responsible for the harms that they knew or should have known are coming from their products. Until there is an incentive to put safety first, safety will remain—at best—an afterthought.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. BRIAN SCHATZ TO HANY FARID

Consumer Right to Identify AI-generated Content

According to a study by the Center for Strategic and International Studies (CSIS), people are only about 50 percent accurate at recognizing content created by AI versus a human—as good as a coin toss. As AI becomes more advanced, it will become even harder, even for tech-savvy people, to tell the difference.

Question 1. In your view, is labeling AI-generated content a viable path to providing this knowledge to people?

Answer. Labelling is necessary but not sufficient. Responsible companies will cooperate with adding content credentials but there will be plenty of services that will not. In this way, labeling could give consumers a false sense of security.

Question 2. Who has the responsibility to step in and establish that transparency for consumers?

Answer. Everyone in the pipeline has responsibility from the content creation to the distribution side.

Question 3. How do you recommend government, AI developers, social media platforms, and others work together to develop those solutions?

Answer. The Coalition for Content Provenance and Authentication (C2PA) and the Content Authenticity Initiative (CAI) are leaders in this space. These organizations have brought together many stakeholders, but many of the largest social media platforms are noticeably absent. The U.S. Government could mandate the labeling of AI-generated content (without passing judgement on the value of this content) and that would, I imagine, be a positive forcing function.

Question 4. In your view, what is the best approach for labeling content generated by open-source AI systems?

Answer. This is a non-trivial problem, and I don't know of any obvious solutions that could not be easily circumvented.

Current Detection Tool Capabilities

AI-generated content detection tools have made significant progress, but are still far from perfect. It's an arms race between those creating increasingly realistic content and those developing the tools to reliably detect it.

Question 1. Will unassisted third-party detection tools ever become accurate enough for consumers to be able to consistently trust their results?

Answer. I don't see an obvious and immediate path to make tools accessible enough or accurate enough for the average consumer. This is why the burden should fall on the AI companies, the social media companies, and the financial sector to deploy technologies to keep us safe.

Question 2. Should developers label their AI-generated content to support consumer trust?

Answer. Yes.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. BEN RAY LUJÁN TO HANY FARID

Question 1. Elaborate on the differences between audio deepfake detection versus visual deepfake detection (image and video). What unique challenges exist for detecting synthetic audio and/or disclosing to consumers that a piece of audio content is synthetic—and in light of the rise in voice-clone scams—what recommendations do you have for policymakers to consider as they work to address the issue of deepfake audio?

Answer. The underlying detection of image, audio, video, is not fundamentally different but the primary challenge with audio in inserting detection in the pipe (phone, voice-mail). This requires the cooperation of the largest telecommunication companies which, frankly, have not been particularly responsive to protecting consumers from a host of annoying and costly scams.

Question 2. Are you aware of any content provenance and authenticity standards efforts for audio content? If so, please describe these initiatives, how they would work in practice, and how scalable this approach would be.

Answer. The efforts of the Coalition for Content Provenance and Authentication (C2PA) and the Content Authenticity Initiative (CAI) operate across all modalities: image, audio, and video.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. MARIA CANTWELL TO JUSTIN BROOKMAN

Liability for Harm

Seniors—and their retirement savings—have long been a target for scammers. AI is making this worse. Recently, an 82-year-old retiree was targeted by scammers who promoted an investment opportunity by using a deepfake video of Elon Musk touting the investment. The consumer opened an account for $248 and ultimately lost his retirement savings—more than $690,000—to the scammers.

Question 1. How do we build safeguards and incentives to prevent similar harms from AI-generated content online? What role should social media platforms play?

Answer. First, we need to do a better job educating consumers about the risks of AI-powered impersonation and the lack of consumer protection and redress mechanisms associated with certain forms of money transfers and investments. People need to become trained to suspect solicitations asking for money in the form of crypto, gift cards, wire transfers, and certain peer-to-peer payment apps, and evoking a false sense of urgency.

However, we should not put all the burden on consumers to protect themselves. We should also update consumer protection laws to provide greater protections for forms of payment beyond credit card transactions—such as payments made through apps instead of credit cards.

We also need to build up and empower regulators such as the FTC, CFPB, and Department of Justice to take speedy action against bad actors and to deter other potential scammers.

We need to develop a transparency infrastructure that embeds provenance data within artificially generated (as well as authentic) content and also requires platforms to surface information to consumers about the nature of the content they host.

Finally, beyond transparency, we should clarify greater responsibilities for toolmakers and online platforms such as social media platforms to take action to limit potential bad uses of generative AI products and to take down fraudulent content (see answer to new question).

Question 2. Who should be held accountable for harmful AI-generated content?

Answer. Certainly, the scammers themselves should be held accountable as the most responsible party, and as noted above, prosecutors need to be given more resources to keep up with the latest generation of digital scammers.

However, in many cases it is difficult to identify or hold accountable the scammers themselves, as they take extreme measures to cloak their identity, and even if identified are often overseas and effectively outside the power of American authorities.

As discussed in my testimony, generative AI tools should have some obligation to take action to remediate potential harms stemming from the use of their products. This is especially the case for tools whose most likely uses are harmful—such as deepfake pornography generators, voice impersonation tools, or the online review generator Rytr that was the subject of a recent FTC enforcement action.[1] In some cases, if the harms created by these tools outweigh the benefits to consumers and competition—and those harms cannot be remediated—those tools should not be made available to the public.

Further, the online platforms that host fraudulent or otherwise harmful AI-generated content should have clearer obligations to take proactive steps to detect and remove such content and to respond to complaints from consumers. These platforms likely already have such obligations under Section 5 of the FTC Act but updating the law could provide better clarity both to companies and to the users of their services.

Privacy and AI-enabled Fraud

When consumers lack control over their personal data, it can be used against them. We already see this with non-AI fraud, where bad actors can use personal details about consumers to perpetuate scams and gain access to consumer account information.

With generative AI, bad actors can use consumers' biometric data, including their voices and faces, against them by creating AI-generated deepfakes. This raises the stakes for protecting consumer data and giving consumers meaningful rights over their data—and the ability to enforce that right in court.

Question 1. How can providing consumers with more rights to control their data help prevent fraud and scams—including AI-enabled fraud and scams?

Answer. If we limit the amount of data about us that is available for purchase on the open market, we can significantly limit the effectiveness of targeted social engineering scams, such as spear phishing attacks. For years, scammers have been able to purchase detailed information about consumers in order to specifically tailor impersonation schemes. Using AI, these same scammers can purchase personal data at scale and use it to craft automated personalized solicitations that make phishing and other attacks significantly more effective.

However, framing privacy protections in terms of "rights" is not the best way to think about privacy. Granting consumers rights puts the onus of protecting their personal information on them, which in practice can be overwhelming. No one has the capacity to micromanage privacy settings across the dozens of businesses with which they regularly interact—let alone the hundreds of different websites and mo-

[1] *In the Matter of Rytr,* LLC, Fed. Trade Comm'n, File No. 232–3052, Complaint, (Sep. 25, 2024), *https://advocacy.consumerreports.org/wp-content/uploads/2022/09/CR-Endorsement-Guides-comments-September-2022–3.pdf;* Consumer Reports filed a comment on the Rytr proceeding in support of the settlement, arguing it was in the public interest and that Rytr's product could only be reasonably used for fraudulent purposes. *See* Justin Brookman *et al., Consumer Reports files comment in support of FTC's settlement with Rytr,* (Nov. 4, 2024), Consumer Reports Advocacy, *https://advocacy.consumerreports.org/research/consumer-reports-files-comment-in-support-of-ftcs-settle ment-with-rytr/.*

bile applications they visit. Professor Dan Solove, among others, have written extensively on the limitations of protecting privacy through the allocation of "rights."[2]

Instead, consumer privacy should be protected by default. For this reason, we have urged Congress and other lawmakers to enact privacy legislation based on the principle of "data minimization."[3] Under a data minimization framework, companies are limited to processing personal information directly in service of a consumer request, with limited statutory exceptions for necessary first-party operational uses such as security and analytics.

At the very least, if consumers are given rights to safeguard their data, they need to be able to exercise these rights at scale, either through dedicated agents working on their behalf,[4] or through universal tools such as the Global Privacy Control (of which Consumer Reports is a founding organization).[5]

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. BRIAN SCHATZ TO JUSTIN BROOKMAN

The Landscape of Information Consumption

I believe a major way to rebuild trust in online content is to require platforms to show proof of where the content came from and how it was made. But some people worry that adding too much information could confuse people instead of helping them, even leading them to ignore or undervalue the labels over time. Users may even over-rely on disclosures and assume that unlabeled content is authentic when it clearly isn't.

Question 1. Do you think labeling AI content can help people trust what they see, or could it make things overwhelming?

Answer. Consumer Reports supports content labeling to help consumers distinguish between authentic and synthetic content. Certainly such labels should not try to convey too much information—the labels themselves should be designed to convey fairly minimal information, though a consumer could hover over or click it in order to receive more information (this is how standardized disclosures generated by the Coalition for Content Provenance and Authenticity (C2PA) work).

We believe that California's AI Transparency Act is a promising start, requiring large generative AI platforms to (1) embed latent provenance data within synthetic media, (2) create tools to scan media for such latent data, and (3) offer creators an option to visibly mark media as synthetic.[6] Ultimately this approach is unlikely to be sufficient on its own, but it does begin to put meaningful transparency obligations on generative platforms rather than leave everything to self-regulation (which has certainly not been sufficient).

Bad actors will seek to avoid transparency requirements—and smaller models, including models outside the reach of U.S. regulators—will have the capacity to make increasingly convincing deepfake content. At the very least, however, through transparency standardization and requirements, good actors may be able to cryptographically certify reliable information about content, and platforms will be empowered to take action to disincentivize content creators who fail to do the same.

Question 2. How can we be smart about not only developing these tools to label and disclose AI-generated content, but also doing so in a user-friendly way?

Answer. We should be careful not to put too much onus on users to decipher complicated disclosures or to be able to independently distinguish between authentic and synthetic media. We also should be careful to message the limitations of labels, especially in early days when adoption may be sporadic even among well-intentioned actors, and platforms are not taking enough responsibility to weed out fraudulent content on their services.

Ultimately generative AI tools and the platforms that host content need to take more responsibility to detect and prevent fraudulent and other harmful uses of AI,

[2] Dan Solove, *The Limitations of Privacy Rights,* 98 Notre Dame Law Review 975 (2023), *https://scholarship.law.gwu.edu/cgi/viewcontent.cgi?params=/context/faculty_publications/article/2856/&path_info=SoloveLimitations_of_Privacy_Rights_FINAL.pdf.*

[3] Consumer Reports and the Electronic Privacy Information Center, *How the FTC Can Mandate Data Minimization Through a Section 5 Unfairness Rulemaking,* Consumer Reports, (Jan. 26, 2022), *https://advocacy.consumerreports.org/wp-content/uploads/2022/01/CR_Epic_FTCDataMinimization_012522_VF_.pdf.*

[4] Consumer Reports offers a free product called Permission Slip that helps users take advantage of new privacy rights created by legislation in several states. *See* Permission Slip, *https://www.permissionslipcr.com/.*

[5] Global Privacy Control, *https://globalprivacycontrol.org/.*

[6] CA SB–942 California AI Transparency Act (2024), *https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB942.*

and to help consumers understand the nature of the content they are interacting with. Regulators and enforcers too will have to be more active in pursuing both scammers themselves but also intermediaries that fail to take reasonable steps to address the harms they enable.

Impact to Vulnerable Groups

Increased access to AI tools has widened its scope of users, including scammers. Financial scams, including real estate investment scams that have recently impacted residents in Hawaii, can use AI impersonations to trick everyday people into handing over money.

Question 1. Could you please share how certain groups, such as seniors and those who speak English as a second language, which are both significant communities in Hawaii, could be especially vulnerable to these scams?

Answer. Scammers certainly prefer to target vulnerable populations. Older Americans are at the top of that list—according to the FBI, last year seniors lost over $3.4 billion to scams, up over eleven percent from the year before.[7] The use of AI voice cloning to impersonate loved ones is colloquially called the "grandparent scam"—as scammers often target grandparents by impersonating the voices and identities of grandchildren in need.[8]

Other vulnerable communities are also especially at risk of scam attempts. According to a CR nationally representative survey from April 2024, among Americans who have personally encountered a cyberattack or digital scam attempt, Black and Hispanic Americans are twice as likely to have lost money to a digital attack or scam as white Americans: Thirty-three percent of Black Americans and 30 percent of Hispanic Americans who have encountered such an attack say they have lost money to a digital attack or scam, compared with just 13 percent of white Americans.[9]

This finding is backed by several other studies, including a 2016 report by the Federal Trade Commission, "Combating Fraud in African American and Latino Communities," which found that communities of color are more likely to become victims of fraud and are more likely to lose money when they are victims of fraud.[10] It is also the case that there are some scammers who specifically target communities who do not primarily speak English. For example, earlier this year the Federal Trade Commission brought an action against Ganadores Online and Ganadores Inversiones Bienes Raíces for targeting Spanish speakers with business opportunity and real estate scams.[11]

Community Resilience Against Scams

I'm proud to have introduced with my colleague Senator Young, the bipartisan AI Public Awareness and Education Campaign Act, which passed out of this Committee earlier this year, to try to get at these concerns about risks of AI in the daily lives of Americans.

Question 1. How would you recommend making vulnerable communities more resilient to identifying and protecting themselves from these scams?

Answer. We need a public awareness campaign to proactively warn consumers about the harms of AI-powered scams and basic precautions that consumers can take to protect themselves. Given how scams are disproportionately targeted at vulnerable consumers, such a campaign should be especially targeted to reach those communities through the channels most likely to reach them.

This campaign should make Americans aware of the increasing risks associated with deepfake and other AI-powered targeted scams (like phishing and robocalls), and teach basic security precautions such as the use of multifactor authentication,

[7] Alanna Durkin Richer, *Scammers stole more than $3.4 billion from older Americans last year, an FBI report says,* Associated Press, (Apr. 30, 2024), *https://apnews.com/article/older-people-fraud-fbi-report-c0da7899f667f9daace4926d5ff3f427.*

[8] *'Grandparent' Scams Get More Sophisticated,* Federal Communications Commission, (Feb. 1, 2024), *https://www.fcc.gov/grandparent-scams-get-more-sophisticated.*

[9] Yael Grauer, *New Report: 2024 Consumer Cyber Readiness,* Consumer Reports Innovation Blog, (Oct. 1, 2024), at 4 *https://innovation.consumerreports.org/new-report-2024-consumer-cyber-readiness/.*

[10] Report to Congress, *Combating Fraud in African American & Latino Communities The FTC's Comprehensive Strategic Plan,* Federal Trade Commission, (Jun. 15, 2016), *https://www.ftc.gov/system/files/documents/reports/combating-fraud-african-american-latino-communities-ftcs-comprehensive-strategic-plan-federal-trade/160615fraudreport.pdf.*

[11] Press Release, *FTC Action Leads to Ban for Ganadores Real Estate and Income Scam, its Owner, and Managers,* Federal Trade Commission, (Jan. 17, 2024), *https://www.ftc.gov/news-events/news/press-releases/2024/01/ftc-action-leads-ban-ganadores-real-estate-income-scam-its-owner-managers.*

promptly installing security updates, how to block third-party content on the internet, and which payments methods offer stronger consumer protections and redress mechanisms (and which do not).

Finally, transparency obligations on generative AI creators, tools, and the digital platforms that host such content—as well as detection and remediation responsibilities for the latter two categories—will help provide actionable information to vulnerable communities and also protect them from being exposed to fraudulent content in the first place.

Response to Written Question Submitted by Hon. Tammy Baldwin to Justin Brookman

Question. Representative Jamie Raskin and I have been leading an effort in the FY25 Appropriations bill to direct the Federal Trade Commission to develop a comprehensive online resource that will serve as a centralized resource page for victims of financial scams and frauds.

During the hearing, I asked you to describe the steps people must take to report the crime and recover damages once people realize they have become the victim of financial fraud or a scam.

What resources do you believe would be the most helpful for victims?

Answer. A centralized resource page should prioritize guidance on potential remedies available to consumers, and provide detailed information about the differing protections for different payment methods. It should also provide guidance on other steps consumers should take based on what information or devices have been compromised, such as how to freeze credit and to remotely disable access to accounts. Consumers should also be informed of how to report illegal activity to the appropriate authorities. The FTC currently offers guidance on most of these topics, though its scams portal would benefit from visual redesign, and greater specificity on the limitations on protections for certain payment methods would be welcome.

In addition to reactive guidance for people who have been harmed by scams, policymakers also need to invest in affirmative education to consumers to prepare them in advance to avoid falling for such scams in the first place. Consumer Reports for example offers Security Planner with practical advice to consumers on how to identify phishing and other social engineering attacks.[12] Unfortunately, we cannot always rely on consumers to seek out static resources; instead there should be a public awareness campaign to prospectively alert consumers to the risk of AI-powered scams. We are currently part of such a campaign called “Take9” that messages to consumers they should pause and take nine seconds to consider whether urgent calls to action are legitimate or not, but a broader effort is needed.[13] Given the staggering consumer losses to scams each year, investing in such a campaign would likely save Americans billions of dollars on net.[14]

Response to Written Questions Submitted by Hon. Ben Ray Luján to Justin Brookman

Question 1. The Coalition for Content Provenance and Authenticity (C2PA) has introduced a content credentials “icon of transparency”, which is open source and can be added to video and image content. Hovering over the icon would reveal a “digital nutrition label” containing information such as: “the publisher or creator’s information, where and when it was created, what tools were used to make it, including whether or not generative AI was used, as well as any edits that were made along the way.”[15] Does Consumer Reports believe that it should be voluntary for a social media company or online platform to display content credentials information for the content that they distribute? Why or why not? Please note that this question is focused on whether the disclosure of content credentials *information* should be voluntary/mandatory, not whether the use of the C2PA’s icon should be voluntary/mandatory.

Answer. CR supports transparency mandates for social media and other platforms to provide access to information about the provenance of media displayed on those services. We also support obligations on generative AI platforms to embed standard-

[12] Security Planner, Consumer Reports, https://securityplanner.consumerreports.org/.

[13] *Nine Seconds for a Safer World,* Take9, *https://pausetake9.org/.*

[14] *Consumer Sentinel Network Data Book 2023,* Federal Trade Commission, (Feb. 2024), *https://www.ftc.gov/system/files/ftc gov/pdf/CSN-Annual-Data-Book-2023.pdf* (noting consumers lost over $10 billion to fraud in 2023).

[15] Introducing Official Content Credentials Icon—C2PA

ized latent information about the provenance of content within synthetic media. While bad actors will seek to evade and frustrate such transparency requirements, at the very least good actors may be able to cryptographically certify reliable information about content, and platforms will be empowered to take action to disincentivize content creators who fail to do the same.

At this point we have not developed a position on whether such disclosures should be in the form of the specific digital nutrition label formulated by the C2PA. We are currently engaging with members of the C2PA and digital platforms to learn about implementation and adoption of C2PA standards and next steps for the coalition's work.

Question 2. If the display of content credentials information were to remain voluntary, what recommendations does Consumer Reports have for consumers who use platforms that do not display this information and are unsure of whether a piece of content is AI-generated? In other words, how can we best equip consumers with the information they need to understand and interpret the media content they are seeing?

Answer. Consumers unfortunately will need to be trained to doubt the authenticity of content they view online, especially through social media and other platforms without a robust editorial process. Today, there are certain recurring indicators of AI generated content that consumers can be inured to look for—such as unlikely misspellings, uncanny latency, and additional limbs. Over time however, it is likely that AI will improve and it will become harder and harder for even trained individuals to visually distinguish between authentic and inauthentic content. Even absent standardized provenance data, however, platforms and user agents like browsers will be incentivized to develop their own AI-empowered tools to detect and identify synthetic or altered media. However, it is unknown how effective such tools would be given the neverending cat-and-mouse game between the developers of tools that create inauthentic content and the developers of tools that try to identify them.

Response to Written Questions Submitted by Hon. Maria Cantwell to Mounir Ibrahim

Public Private Collaboration

AI-enabled fraud does not only impact consumers. Industry—especially the financial services industry—is also impacted. Deloitte's Center for Financial Services predicts that generative AI could lead to $40 billion in fraud losses in the United States in 2027. Public awareness of deepfakes benefits industry and vice versa.

Question 1. How can Congress encourage public and private entities to work together to educate the public about AI-enabled fraud?

Answer. We fully agree with this assessment and recognize that private enterprises, particularly CISA-designated critical industries such as financial services, are—and will increasingly become—primary targets for AI-driven deception and fraud. This threat has a direct and indirect impact on the general public, as the vast majority of consumers depend on financial services, peer-to-peer commerce, and telecommunications in their daily lives.

To address this pressing issue, we encourage collaborative efforts between Congress, the technology sector, and private industry. We propose the following initiatives to enhance public education—targeting both the private sector workforce and the general public:

- *Public Awareness Campaigns:* Congress should consider allocating funding to industry-specific organizations (or agencies) to amplify awareness and spotlight fraud tactics within key sectors such as insurance, banking, lending, and commerce. This support would enable agencies and other stakeholders to better educate their constituents on the evolving methods and techniques used in AI-driven threat vectors, fostering a more informed and resilient approach to combating fraud. These campaigns will help increase efforts from private industry in those categories to address these threat vectors and adjust processes to help mitigate the risks of fraud.
- *Foster Best Practice Sharing:* Congress is uniquely positioned to establish forums, either public or private, where organizations can collaboratively share best practices for addressing and mitigating emerging threats. These forums could be modeled after the AI Insight Forums but with a more focused scope for each session (*i.e.,* supply chain, banking, healthcare, etc. . .), emphasizing the sharing of mitigation strategies, their effectiveness, and insights into new threats impacting critical industries and consumers. A bipartisan initiative involving private industry, state/local government, civil society, and other key

stakeholders could serve as a powerful mechanism to educate and equip industry-leading organizations with the knowledge and tools needed to better protect their consumers.

- *Incentivize Public-Private Partnerships:* Congress could consider establishing grants, tax incentives, or other financial mechanisms to encourage companies to invest in technologies aimed at fraud prevention and consumer protection. While many companies may be aware of these threats, they often lack sufficient motivation to adjust their operating procedures when they have not yet experienced a direct impact. These incentives provide a proactive and constructive approach to addressing emerging risks without imposing new mandates or regulatory burdens.
- *Funds for Government Implementation:* Congress could consider earmarking funds for government agencies to adopt the same fraud prevention and consumer protection technologies it encourages the private sector to utilize. Agencies like the SBA and FEMA, whose programs often mirror private sector operations, are well-positioned to lead by example. By deploying authenticity and transparency technologies, the government can set a standard and serve as a model for industry adoption, demonstrating the effectiveness of these tools in protecting consumers and reducing fraud. Successes can be used as educational tools for industry and the general public and amplify the overall goals.

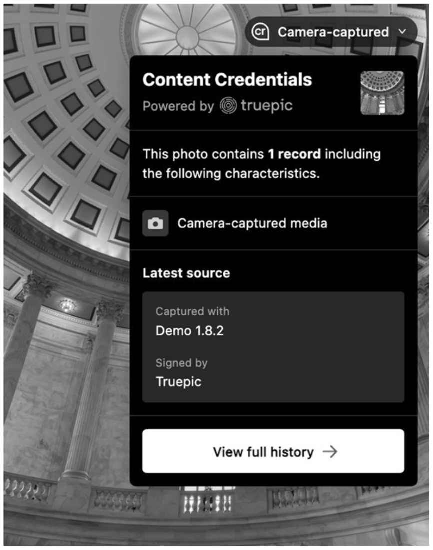
RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. BRIAN SCHATZ TO MOUNIR IBRAHIM

Differences in Content Authenticity vs. Disclosure Approaches

There are two primary approaches for identifying manipulated content. The first is the C2PA route, which is a standard for verifying the source and integrity of the content. The other is to label or disclose synthetic content, such as AI-generated materials. Both approaches have pros and cons: authenticity builds trust but requires more technical systems to verify, while disclosures are more straightforward to the consumer but rely on creators' compliance and are more easily circumvented.

Question 1. What do you think are the advantages and disadvantages of using content authenticity to confirm the source of online material compared to content disclosures that identify synthetic content like AI-generated posts?

Answer. Disclosure of the origins of content can occur with either authentic or synthetic content. The C2PA standard, or Content Credentials, is a form of disclosure when platforms display or "pass through" the transparency information to the content consumer. Below are two examples of using Content Credentials for disclosures: one for authentic and one for synthetic. You can see each clearly labeled as such.



Content Credentials

Powered by truepic

This photo contains **1 record** including the following characteristics.

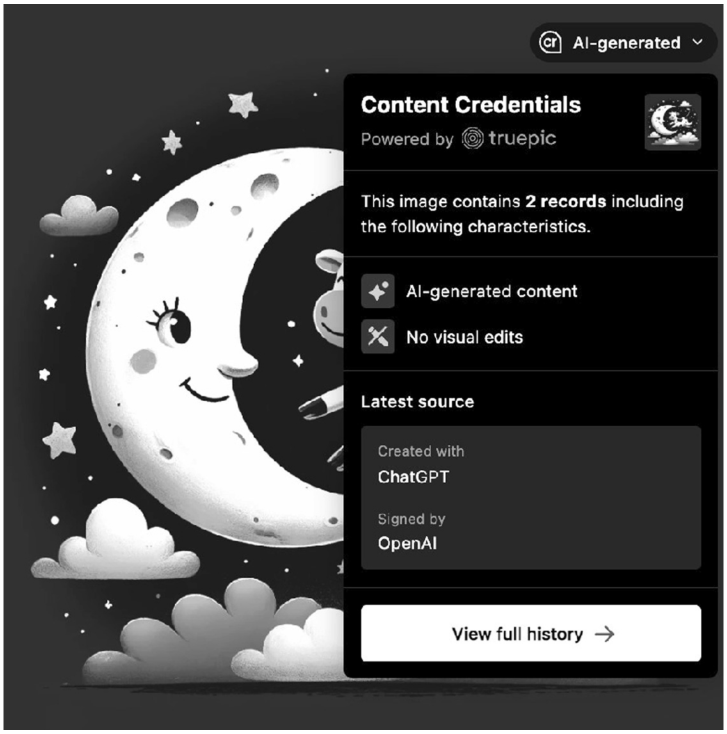
Camera-captured media

Latest source

Captured with
Demo 1.8.2

Signed by
Truepic

Example of authentic image with Content Credential disclosure



Content Credentials

This image contains **2 records** including the following characteristics.

AI-generated content

No visual edits

Latest source

Created with
ChatGPT

Signed by
OpenAI

Example of synthetic image using C2PA Content Credentials to disclose its origins

For either of these images to scale and move through the Internet without losing their cryptographic seal (which would occur if the image were moved to a non-compliant platform) adoption would be necessary throughout the many facets of the internet. This includes OEMs, CDNs, browsers, and distribution platforms. While we are encouraged to see growth and interest in the C2PA's Content Credentials, we also recognize that this process will take time, education, and buy-in with the approach.

We agree that content authenticity extends beyond simply disclosing the origins and creation mechanisms of digital content. A key distinction lies in validating the authenticity of all data and information associated with the content. This requires a robust attestation mechanism to ensure the metadata cryptographically embedded in the file is accurate and verified. The C2PA is actively exploring how creation mechanisms with varying levels of security can be effectively communicated through Content Credentials. Truepic takes this a step further by addressing challenges like rebroadcast attacks, where bad actors capture an image of a screen displaying another image. In such cases, while the metadata may be authentic, the content itself is deceptive. Truepic's expertise lies in preventing such scenarios, ensuring the integrity of both the metadata and the content. Other methods, such as watermarks and digital fingerprints, can also play a vital role in authenticating content by tracing its origin and ensuring it remains untampered. When used in combination, these mechanisms complement each other, creating a more comprehensive and robust framework for digital content authentication. Together, they strengthen the ability to verify authenticity, protect against tampering, and build trust in the integrity of digital media.

Question 2. Do you think these two approaches complement each other or is one preferable to the other?

Answer. We firmly believe that content authenticity and labeling through C2PA Content Credentials are complementary and mutually reinforcing. Truepic's flagship platform, *Vision,* embodies this belief. Vision is a comprehensive content authenticity platform where C2PA Content Credentials serve as a foundational component, augmented by additional veracity and authenticity tests.

These combined efforts ensure our partners and clients receive the highest level of confidence in the integrity of their digital content. Ultimately, we envision the integration of Content Credentials with advanced authenticity mechanisms becoming a best practice, delivering the greatest value to both enterprises and consumers alike.

Response to Written Questions Submitted by Hon. Ben Ray Luján to Mounir Ibrahim

Question 1. The Coalition for Content Provenance and Authenticity (C2PA) has introduced a content credentials "icon of transparency", which is open source and can be added to video and image content. Hovering over the icon would reveal a "digital nutrition label" containing information such as: "the publisher or creator's information, where and when it was created, what tools were used to make it, including whether or not generative AI was used, as well as any edits that were made along the way."[1] Do Truepic and the C2PA believe that it should be voluntary for a social media company or online platform to display content credentials information for the content that they distribute? Why or why not? Please note that this question is focused on whether the disclosure of content credentials *information* should be voluntary/mandatory, not whether the use of the C2PA's icon should be voluntary/mandatory.

Answer. We believe it is essential for any distribution platform—whether a social media platform, commerce platform, or exchange—to effectively relay transparency and authenticity signals embedded in digital content to the end user. When content carries such information, designed to inform or empower consumers to make more informed decisions, it should be conveyed through a reliable and accessible mechanism that preserves its intent and value.

We also acknowledge that the style and method of relaying transparency information to consumers may vary depending on the platform and its user base. Essential details, such as whether the content originates from an AI or synthesis mechanism or was captured by a human using a camera, should always be clearly communicated. However, other metadata—such as time, date, or location—may not always be relevant and could be displayed at the discretion of the platform, tailored to the needs of its audience. Importantly, while the use of the exact C2PA icon and display may not be necessary for every piece of content, we strongly believe that preserving the intent of transparency markings—empowering consumers to make more informed decisions—is critical.

Question 2. If the display of content credentials information were to remain voluntary, what recommendations do Truepic and C2PA have for consumers who use platforms that do not display this information and are unsure of whether a piece of content is AI-generated? In other words, how can we best equip consumers with the information they need to understand and interpret the media content they are seeing?

Answer. On platforms operating under the status quo, where content consumers have no indication of authenticity or synthesis, the best guidance we can offer is to emphasize fundamental media literacy principles. Educating users to critically evaluate content, question sources, and recognize potential manipulation remains essential in such environments. Some publicly available technical tools like reverse image search and EXIF viewers can potentially provide clues and added context on the image but the authenticity of results should always be in question.

Question 3. Are content provenance and authenticity standards efforts underway for audio content? If so, please describe these initiatives, how they would work in practice, and how scalable this approach would be.

Answer. Yes, they are. The C2PA's Content Credentials can be added to audio and Truepic's technology does so. They can be used for authentic audio or synthetic creations alike. Similar to an image the creation mechanism (recording device like a phone or video camera, or synthesis platform) would add Content Credentials at the point of creation. However, it is important to note that audio is a much more challenging medium to ensure authenticity—even with the presence of Content Credentials.

Unlike imagery, preventing rebroadcast audio attacks is incredibly difficult. For example, bad actors can capture audio of another audio while adding Content Credentials to the recapture to deceive consumers into thinking it is authentic. With

[1] Introducing Official Content Credentials Icon—C2PA

imagery, various content authenticity techniques can detect rebroadcast visually, but that same capability does not exist for audio at reasonable accuracy levels.

For these reasons, Truepic only provides its Content Credentials for audio in closed systems, in which it is used as one data point for an enterprise or organization or for the synthetic creation of audio. In which the audio is immediately marked as synthetic from its creation.