S. HRG. 118–573

OVERSIGHT OF AI: ELECTION DEEPFAKES

HEARING

BEFORE THE

SUBCOMMITTEE ON PRIVACY, TECHNOLOGY AND THE LAW

OF THE

COMMITTEE ON THE JUDICIARY UNITED STATES SENATE

ONE HUNDRED EIGHTEENTH CONGRESS

SECOND SESSION

APRIL 16, 2024

Serial No. J–118–63

Printed for the use of the Committee on the Judiciary

www.judiciary.senate.gov
www.govinfo.gov

U.S. GOVERNMENT PUBLISHING OFFICE
58–917 WASHINGTON : 2025

COMMITTEE ON THE JUDICIARY

RICHARD J. DURBIN, Illinois, *Chair*

SHELDON WHITEHOUSE, Rhode Island
AMY KLOBUCHAR, Minnesota
CHRISTOPHER A. COONS, Delaware
RICHARD BLUMENTHAL, Connecticut
MAZIE K. HIRONO, Hawaii
CORY A. BOOKER, New Jersey
ALEX PADILLA, California
JON OSSOFF, Georgia
PETER WELCH, Vermont
LAPHONZA BUTLER, California

LINDSEY O. GRAHAM, South Carolina, *Ranking Member*
CHARLES E. GRASSLEY, Iowa
JOHN CORNYN, Texas
MICHAEL S. LEE, Utah
TED CRUZ, Texas
JOSH HAWLEY, Missouri
TOM COTTON, Arkansas
JOHN KENNEDY, Louisiana
THOM TILLIS, North Carolina
MARSHA BLACKBURN, Tennessee

JOSEPH ZOGBY, *Majority Staff Director*
KATHERINE NIKAS, *Minority Staff Director*

SUBCOMMITTEE ON PRIVACY, TECHNOLOGY AND THE LAW

RICHARD BLUMENTHAL, Connecticut, *Chair*

AMY KLOBUCHAR, Minnesota
CHRISTOPHER A. COONS, Delaware
MAZIE K. HIRONO, Hawaii
ALEX PADILLA, California
JON OSSOFF, Georgia

JOSH HAWLEY, Missouri, *Ranking Member*
JOHN CORNYN, Texas
MICHAEL S. LEE, Utah
JOHN KENNEDY, Louisiana
MARSHA BLACKBURN, Tennessee

DAVID STOOPLER, *Democratic Chief Counsel*
JOHN EHRETT, *Republican Chief Counsel*

CONTENTS

OPENING STATEMENTS

WITNESSES

APPENDIX

OVERSIGHT OF AI: ELECTION DEEPFAKES

TUESDAY, APRIL 16, 2024

UNITED STATES SENATE,
SUBCOMMITTEE ON PRIVACY, TECHNOLOGY,
AND THE LAW,
COMMITTEE ON THE JUDICIARY,
Washington, DC.

The Subcommittee met, pursuant to notice, at 3:33 p.m., in Room 226, Dirksen Senate Office Building, Hon. Richard Blumenthal, Chair of the Subcommittee, presiding.

Present: Senators Blumenthal [presiding], Whitehouse, Klobuchar, Hirono, Padilla, and Hawley.

OPENING STATEMENT OF HON. RICHARD BLUMENTHAL, A U.S. SENATOR FROM THE STATE OF CONNECTICUT

Chair BLUMENTHAL. The hearing of the Subcommittee on Privacy, Technology, and the Law will come to order. Welcome, everyone. I apologize for the delay. Senator Hawley, the Ranking Member, will join us when he arrives. And I want to apologize to our witnesses about the delay. As you know, the full Senate today heard the articles of impeachment from the House, and so we were in our chairs to hear them.

We're here today at this Subcommittee meeting because a deluge of deception, disinformation, and deepfakes are about to descend on the American public. The form of their arrival will be political ads and other forms of disinformation that are made possible by artificial intelligence. There is a clear and present danger to our democracy.

This world of disinformation doesn't have to be our future. We have agency. We can take action. And we are here today not only to hear about the dangers but also to look forward to action that we can take in the U.S. Congress. But we should make no mistake. The threat of political deepfakes is real. It's happening now. It's not science fiction coming at some point in the future, possibly or hypothetically. Artificial intelligence is already being used to interfere with our elections, sowing lies about candidates and suppressing the vote.

We already have a chilling example. This January, thousands of New Hampshire residents received a call impersonating President Biden, telling them not to vote—not to vote—in the State's primary. And it's important for the American people to hear exactly what was said.

[Video is shown.]

Chair BLUMENTHAL. It's important for the American people to hear what impersonation and deepfakes look like. It's also important to know that that's what suppression of voter turnout looks like. The deepfake of President Biden wasn't made by a computer whiz or some computer science graduate student or anybody with any particular skill. It was made by a street magician whose previous claim to fame was that he has world records in spoon bending and escaping straightjackets. The voice-cloning technologies used in that call were inconceivable just a few years ago. Now, they are free, online, available to everyone. And it's not just voice cloning. Deepfake images and videos are disturbingly easy for anyone to create.

Protecting our elections isn't about Democrats versus Republicans. Already, deepfakes have targeted candidates from across the political spectrum, and no one—literally no candidate, no voter—no one is safe from them. And if a street magician can cause this much trouble, imagine what Vladimir Putin or China can do. In fact, they're doing it. National security officials and law enforcement have been shouting from the rooftops, as well as in our classified briefings, their fears about AI and foreign disinformation. It's happening. It's here.

Earlier this month, Microsoft revealed that social media accounts linked to the Chinese Communist Party were using AI to meddle in American politics. China has been caught using deepfakes to impersonate Americans, to sow division and conspiracy theories, such as deepfake images to push the lie that the United States military caused wildfires in Hawaii. Between the ease of use and the increasing interest from foreign adversaries and domestic political interests, our democracy is facing a perfect storm.

When the American people can no longer recognize fact from fiction, quite literally, it will be impossible to have a democracy. As we discussed in our last hearing, these deepfakes and rampant disinformation are also happening at a time when local journalism is hanging by a thread. Deepfakes have targeted not only Presidential candidates but also Senate campaigns and local elections like the recent Chicago mayoral election. Anyone can do it, even in the tiniest race.

In some ways, local elections present an even bigger risk. A deepfake of President Biden will attract national attention. It will be publicized as disinformation and deception, but deepfakes on a local election, State legislative contest, or city council, probably not. And when a local newspaper is closed or understaffed, there may be no one doing fact checking, no one to issue those Pinocchio images, and no one to correct the record. That's a recipe for toxic and destructive politics.

Congress has the power, indeed the obligation, to stop this AI nightmare. There are common-sense, bipartisan bills ready to go, right now. I'm supporting them. A number of my colleagues have offered and supported them, as well: Senators Klobuchar and Hawley's legislation to prohibit deceptive political deepfakes, the Protect Elections from Deceptive AI Act; requiring consent and watermarks for deepfakes, such as Senator Coons and Senator Blackburn—it's called the NO FAKES Act. On Section 230, Senator Hawley and I have a bill to ensure that there's no question that

Section 230 does not apply to AI. We can hold social media companies and Big Tech accountable for election deepfakes and other AI-driven harms.

If we leave them unchecked, deepfakes and political deceptions will sow the seeds of our destruction as a democracy. It may sound like an exaggeration, but it is dangerously true. And so this world of disinformation and poisonous lies doesn't have to be our future. Today, we need to begin or continue the process of making sure it isn't our future. And with that, I'll turn to the Ranking Member.

OPENING STATEMENT OF HON. JOSH HAWLEY, A U.S. SENATOR FROM THE STATE OF MISSOURI

Senator HAWLEY. Thank you very much. Mr. Chairman, thanks for holding this hearing. Thank you for being here. All right. I'm glad we got that on the record.

[Audience disruption.]

Senator HAWLEY. Thanks to the witnesses for being here. I just want to add—I don't really have much to add to what the Chairman said, because I just agree with all of it. I think this issue is—it's not just an issue anymore. It's not just a theory, and the effect of AI on elections is not something—deepfakes in elections is not something that is any longer just a theory. We've seen it. I mean, we've seen it happen. I mean, some of you are here today to testify about it.

We've seen it with fake robocalls. We've seen it with fake images, fake videos produced and disseminated on social media having to do with candidates. It's not confined to one political party or to one primary. It's happened in multiple—all across the country. And I think the dangers of this technology without guardrails and without safety features are becoming painfully, painfully apparent. And I think the question now is, are we going to have to watch some catastrophe unfold?

Already we're watching everyday people have their images stolen, have their likenesses used, commandeered. We're watching folks having their images being taken and being turned into pornographic material. We're watching news anchors have their images ripped off, turned into false information, dubbing, effectively, things that they didn't say.

We're watching the effect on elections. Are we going to have to have further disaster? Are we going to have to have an electoral disaster before Congress realizes, gee, we really should do something to give the public some sense of safety, some sense of certainty that what they're seeing and hearing is actually real or is it, in fact, manufactured? And I think that is a baseline that we're talking about here.

But I want to echo and amplify something the Chairman just said, which is that there are multiple bipartisan bills that are common-sense bills that are ready to go, and I'm proud to have worked on them with everybody sitting on this dais, beginning with the Chairman; Senator Klobuchar, who has worked very hard on this. It's time that these bills got a vote. I mean, we can talk and talk, and nobody has done a better job of surfacing this issue and bringing facts into the public domain than the Chairman has, but now it's really time to vote. And I just call on the leadership of both par-

ties in the Senate. Both parties. The leadership needs to support an effort to get a vote.

And I say, an effort to get. Really, they just need to schedule a vote. Let's put these bills on the floor, and let's vote. Let's not allow these same companies that control the social media technology in this country, that control the news in this country, to also now use AI to further their hammer hold on the United States of America and on our political process. So, thank you, Mr. Chairman, again. Thanks to all the witnesses.

Chair BLUMENTHAL. Thanks, Senator Hawley, and thanks for your work on this issue. And I want to turn to Senator Klobuchar, who's really been a leader on this Committee, the Judiciary Committee, but she also chairs the Rules Committee, which will oversee a lot of this legislation when hopefully it does get to the floor. And certainly we're here because we believe there should be a vote. And thanks to Senator Klobuchar for your leadership.

OPENING STATEMENT OF HON. AMY KLOBUCHAR, A U.S. SENATOR FROM THE STATE OF MINNESOTA

Senator KLOBUCHAR. Thank you. Well, thank you, Chairman, and thank you, Ranking Member Hawley, for this important hearing and this opportunity to keep this on the front burner. As Senator Hawley just said, we cannot wait. We are scheduling a mark-up of our bill, and we are going to have to work—it's the only Committee that both leaders are on. Fun Committee to Chair. And so I will seek Senator Hawley's help and others' on our bill—which includes Senator Coons and Collins, Senators Bennet and Ricketts and a whole lot more support on both sides of the aisle—to get the votes, not just to—you know, we can obviously pass it, but I'd like to get a really strong vote coming out of Committee, so we can immediately get this thing heard, because we really can't wait. The elections are upon us.

And like any emerging technology, AI has great opportunities but also significant risks. And this is the one right before us, as well as other issues related to scams, and we have to put rules in place. And we can't let the same thing happen as—every one of the four of us has been out front on this—as happened with Section 230 and what happened when they just acted like these companies were little things in a garage, and now they're humongous monopolies, and now we are all challenged in trying to get these bills forward, whether it's on fentanyl, whether it is on child pornography, whether it is on competition policy. And we have to move these.

The fake robocall—I hadn't actually heard it myself, so thank you for that. And it is just impossible to tell that that's not Joe Biden, as it was impossible to tell one video that ran during the Republican primary that wasn't accurate, involving Donald Trump. That also wasn't accurate. We had an Elizabeth Warren video in which she says Republicans shouldn't vote. That wasn't her, but you couldn't tell.

We had a Minnesota—and this is not AI, but it just shows how devastating this can be. We had a photo, the day after the heroes, the two police officers in Burnsville, Minnesota were killed after rescuing seven kids, and then the paramedic was killed who was performing CPR. A photo of an actual rally picture from 2022, that

I was kind of in the background on, started going around. At the same time, there's some kind of Russian photo going around, saying that I fund Nazis in Ukraine. That's been going around for 3 years.

This photo had a red circle around me in the background, and then they put Defund the Police signs in the hands of the people at the rally, that were never there. So, they were literally using—the people who did this—I personally think it was foreign interests, but—took a photo and put those Defund the Police signs after these officers had been killed. And to their credit, X and Meta put, Altered Content, with a big sign, but it took us about, you know, a day to get all this down. It was going all around the internet. That is actually not AI. That's a real photo that they doctored, and people thought it was real. It looked real.

And so this kind of thing is just going to keep happening and keep happening unless we take immediate, immediate action. Eleven States, including my own, have enacted laws to address these threats to our elections, and that's great, but it doesn't cover Federal. And some of these States are—they're not all blue States, whatever, purple States. People are taking action, as seen by the bipartisan nature of our legislation. And we also need disclaimers on other ads that aren't deepfakes, and that's a bill that we will also be marking up in the Rules Committee.

So, I want to thank my colleagues for doing this. I want to thank them for their willingness to stand up on this issue and look forward to hearing the testimony of the witnesses. Thank you.

Chair BLUMENTHAL. Thanks, Senator Klobuchar. I will now introduce our witnesses. This panel is extraordinarily distinguished. Zohaib Ahmed is the CEO and founder of Resemble AI, a research and development lab focused on the creation of generative voice models. He and his team have spent the last 5 years developing and researching AI voice and detection technology and are uniquely positioned to understand both the remarkable potential and possible risks associated with the rapid advancement of voice synthesis and cloning capabilities.

Rijul Gupta is the visionary founder and CEO of Deep Media, a leading deepfake detection and AI security company, with a foundation in machine learning from Yale University and over 15 years' experience writing AI algorithms. He has dedicated his life to developing Deep Media's patented AI technologies and establishing the company as the gold standard in combating threats posed by unethical AI and deepfake misinformation.

Ben Colman is the cofounder and CEO of Reality Defender, a cybersecurity company helping enterprises and governments detect deepfakes, and data science, for over 15 years. He has had 10 years at Goldman Sachs and Google.

David Scanlan became New Hampshire secretary of State in January 2022 after serving 20 years as deputy secretary of State. Prior to that, he served eight terms in the New Hampshire House of Representatives, including a term as majority leader before joining the secretary of State's office. As is our custom, I'm going to ask you to stand and be sworn.

[Witnesses are sworn in.]

Chair BLUMENTHAL. Thank you. Why don't we go down the panel, beginning with you, Mr. Ahmed.

STATEMENT OF ZOHAIB AHMED, CEO AND CO-FOUNDER, RESEMBLE AI, SANTA CLARA, CALIFORNIA

Mr. AHMED. Absolutely. Thank you. Chairman Blumenthal, Ranking Member Hawley, Members of the Committee, thank you for the opportunity to discuss the oversight of AI as it relates to understanding the impact this technology can have on the election. As I was introduced, Resemble AI is a research and development lab focused on the creation of generative voice AI. We've worked with large media companies, game studios, telecom companies, as well as content creators, to produce AI voices. And we've spent the last 5 years developing and researching this voice technology.

We've created terabytes' worth of data sets, and we're uniquely positioned to understand the remarkable potential and the possible risks. And over the last 9 months, a lot of the research we've opened up—regarding responsible voice cloning technology, including research on speaker identification, watermarking, and deepfake detection. In my testimony, I want to share some of the technologies that we've developed since Resemble AI was founded, especially around watermarking and deepfake detection, and share some of the recommendations I might have around transparency and disclosure, safeguards and mitigation, and integrity verification.

I'd like to pull up a couple of slides just to help the audience just understand. Pull them up. Great. Sounds great. So, before I jump into any of these audio clips that you'll hear in a few seconds, I want to walk through how some of these AI voices are created. I think it's very important to understand how the technology works. And, you know, we take a few minutes, seconds of audio. Like Chairman Blumenthal said, it's super easy to create some of these voices, and these models have become widely accessible at this point.

We've always held ourselves to exceptional standards of ethics, and we've developed many guardrails since the inception of the company, to make sure that our technology is used safely, the first of which is a built-in speaker identification model, which we have open sourced about 2½ years ago. We actually use it internally, to make sure that we get consent from every speaker that uses our technology, so there's no way anyone can basically go in, upload any seconds of audio, minutes of audio, and create voices from there. We also have clear terms—what you can and cannot use the AI voices for.

So, I'd like to play some of these audio clips in the presentation itself. Maybe we can start with the first one on the left-hand side. One of these audio clips that you hear before is real. So, we'll go ahead.

[Audio clip.]

Mr. AHMED. The second one, there.

[Audio clip.]

Mr. AHMED. Go for a third.

[Audio clip.]

Mr. AHMED. And the last one.

[Audio clip.]

Mr. AHMED. So, hopefully you can take a guess at which one's real. You can do it in your heads right now. We'll go to the next slide.

[Audio clip.]

Mr. AHMED. I think it's—there we go.

[Audio clip.]

Mr. AHMED. There we go. The second one is real. So, I'm not sure how many of you got that right. We can move on to the next slide again. If you guessed incorrectly, you wouldn't be the only one. You know, as Senator Blumenthal mentioned, these voice fakes—you know, the Biden one—we've heard him so much that, you know, you know what he sounds like. You can pick up on nuances. This was my colleague, you know, so this was—these were all generated—well, three of them were generated. One of them was real, as you saw. And as you're all aware, this is happening in much more frequency right now.

We acknowledge that the consumer education and awareness is a critical piece of addressing the situation. For the last 12 months, we've been publishing detailed incident reports of every case where AI is utilized for scams. We have analyzed the Joe Biden incident. Yesterday, we analyzed the LastPass CEO that was used as a deepfake incident on WhatsApp. So, you have enterprises; you have consumers. You're all being targeted by the wide spread of technology. And this is all available for anyone on our blog.

Go over to the next slide. After creating and open sourcing the speaker identification model, we then worked to create a Neural Speech Watermarker, as well as a deepfake detection model that has 98 percent accuracy. We've found that this is so critical that we've actually made the deepfake detection tool absolutely free. You can go to detect.resemble.ai, and anyone can drag and drop any file, point to any YouTube link, and figure out whether it's fake or real. We've also integrated into tools like Google Meet, making it widely accessible.

I want to jump to some recommendations, really quickly, here. First and foremost, we support the proposed legislation that requires clear labeling of AI-generated content. To take it one step further, we propose there be a creation of a public data base where all AI—where all generated election content is registered, allowing voters to easily access information about the origin and nature of the content that they encounter. This includes the deepfakes that may be out there.

To adequately safeguard against misinformation, particularly during critical events like election, collaboration is key. By distributing and allowing platforms—or enforcing platforms to use watermarking technology or using deepfake detection will instantly tell the consumer whether something is real or fake. You won't have instances where you have to wait a whole delay while things propagate throughout the world, and then you realize, oh, you have to, like, create community notes, to backstep. We believe that AI watermarking technology is a readily available solution that can already check the integrity of audio content.

We propose that all election-related audio content, including political advertisements, campaign messages, and public statements

by candidates be watermarked with the technology. One of the key aspects of our watermarking is that it can actually persist through training. So, generative models, when they scrape data and train models—we can actually figure out, from the output of the model, where the data came from, which is significantly important. The traceability aspect is really important.

We also recommend the establishment of a certification program, much like you have the check marks and you have e-signatures. Setting standards for the effectiveness and reliability of watermarking solutions ensures that only trusted and vetted technologies are used.

We're always willing to help facilitate partnerships between private and public sectors, to ensure today's innovation is used responsibly. Thank you for the opportunity to provide insight into voice cloning technology and preventative measures that can be taken now to ensure the integrity of this year's election. Thank you.

[The prepared statement of Mr. Ahmed appears as a submission for the record.]

Chair BLUMENTHAL. Excellent. Thank you very much. Mr. Colman. And you can turn on your——

Mr. COLMAN. Thank you, Senator.

Chair BLUMENTHAL. There you go.

Mr. COLMAN. Ahead of my 5 minutes, we had a chance to work with Senator Blumenthal's office to record a few real and fake audio clips, which I'd like to play for the group. If we can start with the first one, we're going to ask the audience and those on the dais which ones are real and which ones are fake.

[Audio clip.]

Mr. COLMAN. And the second one, please.

[Audio clip.]

Mr. COLMAN. And the third and final one.

[Audio clip.]

Mr. COLMAN. And as you guys think about which ones are real and fake, we're going to share with you the surprise that they're actually all fake. And really, the challenge and opportunity is that anybody with a Google search and internet connection can make something as entertaining or as dangerous as they can imagine.

Chair BLUMENTHAL. Anybody ought to know I'm not a Royals fan.

[Laughter.]

Mr. COLMAN. No comment on that.

Chair BLUMENTHAL. With all due respect to the Ranking Member.

STATEMENT OF BEN COLMAN, CEO AND CO-FOUNDER, REALITY DEFENDER, NEW YORK, NEW YORK

Mr. COLMAN. Chairman Blumenthal, Ranking Member Hawley, Senator Klobuchar, and Committee Members, I thank you for your stated concerns on the impacts deepfakes have had on our elections and our democracy, and I thank you for holding this hearing as well as requesting my presence here. It is an honor to provide insights that may help both your Committee and the American peo-

ple, and I applaud the Committee's efforts in surfacing this problem in front of the Nation.

For 3 years, led by unmatched innovation in American tech companies, rapid advancements in generative artificial intelligence are now a permanent fixture in society. As a long-time cybersecurity professional, myself and my team at Reality Defender foresaw the harms these technologies could bring, years before the current AI boom. We built our company because, after seeing how weaponized content and disinformation impacted our loved ones, we sought to combat the future technological drivers of advanced disinformations, which are called deepfakes.

Now, deepfakes are AI-manipulated media that impersonate our citizens or create synthetic identities to spread disinformation or commit fraud. They hit the heart of what makes us human, realistic enough that even those of us who've studied AI for years and have PhDs have been at times unable to tell the difference between real and fake with our naked eyes.

Now, not all AI technology is bad. Now, while they have their benefits, they can also hit core tenets we hold dear. We've seen foreign adversaries wield deepfakes in sophisticated disinformation campaigns, with Russian media falsely depicting Ukrainian forces as the perpetrators of the devastating attack in the Moscow music hall. We saw this in America, with a robocall of a fake President Biden to thousands of New Hampshire constituents, asking them not to vote.

I cannot list every malicious and damaging use of deepfake of the past, present, and future. What I can do is sound the alarm on the impacts they have, not just on democracy but also on America. Anybody with internet access can create AI-generated audio, video, images, or text to convince and persuade millions of people. This fake media can be distributed and shared instantly over social media platforms. The more incendiary the content, the faster it spreads. Trust and safety teams at these platforms once blocked misinformation and fraud from spreading, but now the teams barely exist, leaving the onus of detection and verification on the users.

Ahead of our 2024 election, also a year where two-thirds of the world will be voting in similar elections, we've seen the blueprints of deepfake-fueled interference in Taiwan and Slovenia's most recent elections. In these cases, materials appeared and instantly spread to millions. The responses pinpointing them as deepfakes took substantially longer to spread. By the time the deepfake widely spreads, any report calling it a deepfake is also too late. Uncovering truth will always be slower and harder than spreading a lie.

The same type of deepfake-enabled operations can and have happened here. There will continue to be more damaging results as deepfake technology catapults ahead. This is not fearmongering, AI alarmism, doomerism, or conspiracy-minded hyperbole. It is simply the logical progression of the weaponization of deepfakes.

To protect our democracy and the media that drives it, legislation must mandate that content platforms are responsible for the urgent detection and removal of dangerous deepfakes and AI-generated media. I applaud Members of this Committee on their Protect Elections from Deceptive AI Act. Unlike measures that have more or less given the pen to the largest content and social plat-

forms, this law has a great start. We can go further by imposing real penalties on bad actors using deepfakes to morph reality and on the platforms that fail to stop their spread.

Federal laws should outline penalties specific to the severity of using deepfakes and election disinformation crimes, as the State of Minnesota has done. AI developments move fast. Legislation must move faster, forecasting and potentially anticipating the rapid improvements in the quality and application of deepfakes, all built by companies who move fast and break things. The things here are aspects of society everyone in this room holds dear: democracy, truth, trusting your eyes and your ears. It's not a stretch to say that these are at stake, when anyone can instantly create a deepfake to convince millions of people that they're anybody, saying anything.

We must treat deepfakes with equal or greater importance than the worst kinds of content that existed before it, precisely because it gets to the heart of what makes us human. We must act quickly, or we will be taken by surprise by new attacks on democracy, on elections, and on the very concept of truth.

[The prepared statement of Mr. Colman appears as a submission for the record.]

Chair BLUMENTHAL. Thanks, Mr. Colman. Mr. Gupta.

STATEMENT OF RIJUL GUPTA, CEO, DEEP MEDIA, OAKLAND, CALIFORNIA

Mr. GUPTA. Senator Blumenthal, Senator Hawley, Senator Hirono, Senator Padilla, thank you for having me here. I am truly humbled to be here in front of you. My name is Rijul Gupta. I was born and raised in a small town in Oklahoma. I started building apps and websites when I was 10 years old. I'm a hacker at heart but an entrepreneur by trade.

I started building machine learning applications when I was just 15. I went to Yale, where I studied machine learning academically, and after graduating, after a couple of years, I started reading papers about what we now call generative AI. In 2017, I founded Deep Media because I knew deepfakes were coming, and I committed my life, in that moment, to solving the deepfake problem.

Ever since then, we have worked tirelessly to make sure that people have technology to solve this problem, but first, before getting into that, I think it's important that we define what a deepfake is. A deepfake is a synthetically manipulated, AI-manipulated image, audio, or video that can be used to harm or mislead. This does not cover text, right. Whether you believe that human beings evolved over time or whether we were designed this way, the human mind is hijacked by image, audio, video, and that type of synthetic media content really has the potential to completely dismantle society.

I'm not going to go into too much tech detail, but as legislators, if you're going to legislate medicine, you need to know the difference between Tylenol and Tamiflu, right. So, I want you to keep three terms in your mind when you're talking about this technology. The first is transformer. It is a type of architecture. The second is a generative adversarial network, a GAN. And then the third is a diffusion model. Those three fundamental technologies are what generative AI is about. That covers about 90 percent of

it. All of these models require massive amounts of compute resources and massive amounts of data. We're talking about millions of identities, here. So, just keep that in your heads when you're thinking about this technology.

We've all talked about how deepfakes are coming and how they're basically here, but one thing that is, I think, hard for most people to understand, just intuitively, is scale, right. These deepfakes and these AIs—they're getting really, really good, really fast, right. The quality is basically perfect now. They're getting really cheap to produce. Right now, it's about ten cents per minute, for video. That's going down to one cent, really, really quickly. And the amounts of content, the percentage of content that is on online platforms is approaching as much as 90 percent by 2030. Right? So, we all know the harms. It's important you know the scale of these harms.

Now, we've already seen them impact the elections, right. We have the deepfakes of President Biden announcing the draft, the deepfakes of President Trump getting arrested, the deepfakes of Hillary Clinton endorsing Governor DeSantis, right. All of those are about political assassination. We are also seeing deepfakes be used to create groundswell support, right. The deepfakes of President Trump with Black voters. So, it's important to know that these deepfakes are going to be used for political assassination but also for the opposite, to make politicians seem more relatable.

But I think a bigger threat is actually not the fake content; it's what the fake content does to the real content. Right? When anything could be fake, you don't know what's real anymore. And so we're going to start seeing plausible deniability come into play here, where politicians or anyone in business or anyone at all could just claim an image, audio, video is a deepfake. And that is fundamentally dangerous. People think that AI is going to be like The Terminator. It is much more likely to create a society like 1984. That's what we need to be worried about when we're talking about deepfakes.

But in Silicon Valley, we like to take a solutions approach. So, I am here to tell you today that solutions to this exist. But they need to have buy-in from government stakeholders, generative AI companies, the platforms, investigative journalists—even local journalists—and deepfake detection companies themselves. Those five groups of people need to work together to solve this problem.

I am proud to say that we have helped people like Donie O'Sullivan at CNN, Geoff Fowler at Washington Post, and Amanda Florian at Forbes detect and report on deepfakes. We are members of the WITNESS organization, an independent group that surfaces deepfake detection to reporters. We are part of the DARPA SemaFor, an AI Force program that brings in researchers, corporations, and government resources to solve this problem. We are part of the Content Authority Initiative, alongside companies like Adobe that try to label real content and fake content. We also have several of our own committees that we're leading that bring in the deepfake generative AI folks to label their content, people in research for detection, and Big Tech platforms to adopt this technology to keep people safe.

I'm a believer in the free market. I fundamentally think AI can be used for good. I believe deepfakes represent a market failure. They represent a tragedy of the commons and that this fraud and misinformation is a negative externality and that if we legislate this properly, we can internalize that negative externality and make the AI ecosystem flourish. And with that in mind, I would like to take just a couple of minutes to show you how we can solve the deepfake problem.

So, I have a couple of slides that I'd like to show you. And I want you to get in the mindset of: how does an AI see media? Right? That's kind of what we think about. We try and look through the AI's eyes, in order to detect it. So, again, if we can show the slides here on the screen? Go to the next one, please.

Here are some examples of what our system looks like. Again, we are mapping on the left, there, the proliferation of deepfakes over time, as well as the cost to society if we don't solve this problem. This is costs for fraud, misinformation, and other crimes, right. However, our platform can deliver solutions at scale across image, audio, and video. Next slide, please.

[Slide presentation is shown.]

Here we see some examples of real content and fake content. Again, it's not just about detecting a deepfake, right. It's being able to detect a deepfake while not saying a real thing is fake. Right? That's critical. So, our false positive and our false negative rate is very, very low. And if we have a little bit of time, I'd like to show you just how an AI sees audio. We have some images up here, but on the next slide we're going to see how an AI sees audio.

And this is actually a real piece of audio. Here, that yellow and blue graph—that's what an AI sees. When it's seeing a person's voice and trying to learn from it, it's seeing that. And this is an example of our detectors picking this up as real. And you can fast-forward through this. I don't want to take up too much time. But if you want to go to the final slide, that's an actual real political deepfake.

[Audio clip.]

Fast-forward this one.

[Audio clip.]

This is a real video. We've picked it up as real. Right? And the AI is tracking the face; it's picking up certain key points on a person's face. And if you go to the final slide, here?

[Audio clip.]

This is a deepfake that was produced recently, and maybe we can just play the whole thing.

[Video is shown.]

This is the highest quality deepfake made to date. It is using detection of generative models that aren't released publicly, that use their own detection—or, sorry, generation models that they created, hyper high quality, and we picked it up, right. So, it's about staying on top. At Deep Media, we are both the cat and the mouse in the cat-and-mouse game. We have generative AI technology, but we don't give it out to people. We keep it internally and use that to train our detectors, and that is why we are setting the gold standard.

So, again, honored to be here and happy to answer any questions. You all are the policy folks, and I am here to provide as much information as possible about what the solutions, from a technical standpoint, actually are. Thank you.

[The prepared statement of Mr. Gupta appears as a submission for the record.]

Chair BLUMENTHAL. Thank you, Mr. Gupta. Secretary of State Scanlan.

STATEMENT OF HON. DAVID SCANLAN, SECRETARY OF STATE, STATE OF NEW HAMPSHIRE, CONCORD, NEW HAMPSHIRE

Mr. SCANLAN. Thank you, Chairman Blumenthal and Ranking Member Hawley and Senate Members, for the invitation to be here today. Actually, Senator Padilla, it's great to see a former secretary of State here on the Committee, as well.

On the weekend before January 23, when New Hampshire held its first-in-the-Nation Presidential primary, everything was going very smoothly. The candidates were out doing their last-minute campaigning. All of the polling places were set up and ready to go. They had plenty of ballots and, in typical New Hampshire fashion, we were ready to conduct a really good election.

Weekend went fine, and all of a sudden, on Sunday, I started getting some phone calls from reporters, asking if I knew anything about a robocall that was taking place with President Biden. I went to bed that night wondering what was up. First thing in the morning, we conferred with the attorney general's office, and it was apparent that there was a robocall using AI with President Biden's voice on it, asking individuals not to vote in the election because for Democrats their vote was more important to support him in the general election.

Interestingly, the robocall was spoofed—and I understand that's a term where a call is assigned to somebody else's phone number—to a prominent Democrat in the State of New Hampshire who was a former State party chair and a former member of the Democratic National Committee. Because her phone was associated with the robocall, she's starting getting calls from acquaintances, asking her to clarify what was being asked in the robocall. She very quickly figured out what was happening and reported the incident to the attorney general's office, and they opened up an investigation.

Fortunately, when there is a major election taking place in New Hampshire, the media—both State and national media—are on top of it. They are looking for something to report, especially when things are running very, very smoothly. And so when this surfaced, they jumped all over it, which was actually an opportunity for my office, the attorney general's office, and the Governor's office to inform voters of what was occurring, let them know that what was being said on the robocall was a form of voter suppression, that it was illegal, and that in that specific instance, they should ignore it and make sure that they participate in the election.

And every indication is that they did. New Hampshire had a record turnout in both the Republican primary—but also in the Democratic primary, when you have an incumbent President running for a second term. New Hampshire broke a record in the turnout. So, it is hard to tell how much of an impact that particular

robocall using AI actually had on the voting population. We estimate, or the attorney general's office estimated, through the investigation that they have done to date, that there were between 5 and 25,000 calls made in the State of New Hampshire to voters with that information.

So, clearly, you know, it did not seem to have an impact on that election. In hindsight, though, looking back, the call itself was kind of primitive. And, you know, it is something that could have been done with an impressionist, somebody that could—a real, live person that could imitate the voice of the President in this case and could have done the same thing with a robocall. What was concerning was the ease of which a random member of the public that really doesn't have a lot of experience in AI and technology was able to create the call itself. And I think that if you add what happened with video to go along with that—and we saw some great examples here—you could show candidates in compromising situations that never existed. It could be a State election official giving misinformation about elections, and worse. And, to me, that is incredibly problematic.

Now, I know—you know, I know that there are instances where there's parody and there's humor, and I've seen AI with prominent politicians doing funny things, and it is funny, but it's also quite obvious. I think we have to get a handle on when an AI in elections is intentionally deceptive and malicious. We need to be able to recognize it, stop it, and prosecute it.

[The prepared statement of Mr. Scanlan appears as a submission for the record.]

Chair BLUMENTHAL. Thank you very much. Thanks to all of our witnesses for your really excellent testimony.

Mr. Scanlan, you hypothesize—because we can't know for sure—that the Biden deepfake had minimal impact, but we can't be certain what the vote would have been, but for those calls. And I understand there is an investigation ongoing; the attorney general is conducting it. It's under New Hampshire law. I assume it's criminal law, as well as civil, but there're no Federal remedies. In your view, would it be helpful to have criminal penalties under Federal law specifically aimed at this kind of deception? And I think it was Mr. Colman suggested that criminal penalties could be an effective deterrent, but they have to be really more specific and stringent than they are now.

Mr. SCANLAN. Mr. Chairman, I have to agree that we truly don't know what the impact was on the New Hampshire Presidential primary. We only know that we had a good turnout, and the results were what they were. And we still have an active prosecution going on. The AG in New Hampshire has identified a company or companies that participated and an individual that is a suspect, and they're moving forward with that.

At some point, I believe that there is a Federal component to this because it's going to be a national problem. And I'd like to give a shout-out to Cait Conley, who works with Jen Easterly at CISA. Cait was in New Hampshire on the day of the Presidential primary, and she traveled around to polling places with me, to try and get a handle on, you know, how big this thing actually was, even though that was difficult to determine.

But, yes, I think that, you know, these things, in a national election, are going to be generated nationally, whether it's foreign actors or some other malicious circumstance, and I think we need uniformity and the power of the Federal Government to help put the brakes on that. Instances that happen locally—certainly, Federal Government assistance would be helpful, but I think that should remain the prerogative of State law enforcement and the attorney general.

Chair BLUMENTHAL. With assistance from Federal authorities, where it's appropriate. Let me ask you and the other witnesses—Senator Hawley and I have proposed a framework which includes an independent oversight entity; a set of standards that would be imposed by that entity; a requirement for some licensing before models were deployed; testing to assure that they were safe and effective, just as the FDA reviews drugs to make sure they are safe and effective; and, potentially, penalties such as we've been discussing; as well as export controls, to assure that our national security is protected. Just for the sake of speed, I'm assuming that all of you would agree that some kind of framework like that one makes sense.

Mr. GUPTA. I actually have specific thoughts on that framework. I think it's a good start, but I really think it's important that whatever framework we set adopts what's called a defense-in-depth approach, right. So, we need metadata; watermarking; cryptographic hashing, which is a little complicated, but it's invisible watermarks and a hash data base, kind of like NCMEC; AI detection; and AI poisoning. It also needs to cover both the generative AI platforms and the online platforms. We need both of those folks. We can't just say licensed generative AI companies and leave it at that. Honestly, we need government buy-in, generative AI buy-in, platform buy-in, journalists' buy-in, and then detection companies.

Chair BLUMENTHAL. And all of those points are encompassed by our framework, particularly the watermarking, which——

Mr. GUPTA. Yes. I really think watermarking is getting a lot of attention here, and it really doesn't solve that much of a problem. You need cryptographic hashing: invisible watermarking. That's really important.

Chair BLUMENTHAL. Mr. Colman.

Mr. COLMAN. Yes, just to add onto that—and I think just to unpack two things, here—we're talking about watermarking and cryptographic hashing, effectively what's called provenance. It's either there or it's not. The challenge of that is it presupposes that everybody's going to follow the same rules; all the bad actors will follow the same rules. And, you know, we've seen time and time again, a lot of the applications—whether they're on your phone, in the app store, or online, or they're open sourced—they just aren't going to follow the rules. So, we can't expect everyone to say, hey, we're going to play nice within this walled garden, when the bad actors, by definition, are not playing by the rules at all.

And so, you know, with Reality Defender, we focused on inference. We don't touch any watermarking; we don't touch any personal data. We actually assume we'll never see the ground truth. We'll never even know if this is real or not, which means, instead of saying yes or no, we're taking a more measured, probabilistic ap-

proach—a probability, saying, maybe we're 95 percent confident; maybe we're 62 percent confident. We build that into a larger framework of just one signal among many, to make a better insight to have a platform or a team decide to block or flag a piece of media or a person or an action.

Chair BLUMENTHAL. We're going to adhere to 5-minute rounds on the first round. I hope to come back to this line of questioning, and I apologize that others of you—Mr. Ahmed, you may have some comments, as well, but in deference to my colleagues who have other commitments, I'm going to turn to the Ranking Member.

Senator HAWLEY. Thank you again, Mr. Chairman, and thanks to everybody for being here. Mr. Colman, you raised in your opening statement what is I think my nightmare scenario, which is—you made the comment that pretty soon it's going to be anybody with an internet connection is going to be able to access and use deepfake technology. I wonder if we're there already, though.

I mean, I'm looking at this article from the New York Times just a few days ago. The headline is, Teen Girls Confront an Epidemic of Deepfake Nudes in Schools. The details of this are just unbelievable. I mean, this is a young girl, Francesca Mani, at a high school in Westfield, New Jersey, and she's a tenth grader. All of this is in the article. She's a tenth grader, and she found out that a number of boys in her class had used artificial intelligence software to fabricate sexually explicit images of her and a number of her friends and then were circulating them online and, you know, showing them to classmates—but putting them onto platforms.

Now, I presume that these teenaged boys didn't pay a lot of money in order to do this. In fact, the article goes on to say that they used widely available nudification apps to create these fake photos. So, they take the photos of their classmates from, you know, Instagram or wherever and use those, feed them into this app, and then here you are. And it probably cost them almost nothing.

So, I guess my question to you is, are we at the point now with this technology where we're going to see a flood of AI-generated CSAM, a flood of other sexually explicit material created of adults or young adults? I mean, is this the point that we're at now?

Mr. COLMAN. Ranking Member Hawley, we were at that point 6 months ago. And the challenge for us right now is, where the U.S. is leading the development of a lot of these novel technological tools, we're not leading in regulations to protect from these tools. We have Taiwan, we have Singapore, we even have China with more advanced regulations in the space. And to your point beyond elections, thinking about different types of kind of equally or more dangerous risks from deepfakes, there was recently a House Oversight and Accountability Committee referencing very scary statistics that 98 percent of all deepfakes are actually pornographic; 99 percent——

Senator HAWLEY. I'm sorry. What was the percentage?

Mr. COLMAN. 98 percent——

Senator HAWLEY. 98 percent.

Mr. COLMAN [continuing]. Of all deepfakes. 99 percent of people targeted in deepfakes are women. The 40 most popular deepfake pornography websites have over 143,000 deepfakes—pornography,

unpermissioned—just in the last year getting over 4 billion views. Now, these two numbers are more than the previous 10 years combined. So, when I say this is already a problem—it's been a problem. We're waking up to it now, and the election is just one risk that the larger world of regulations can solve for.

Senator HAWLEY. So, I guess my question is, I mean, given that, what is the most effective regulatory avenues to pursue? I mean, how are we going to empower people like Francesca and her parents and the hundreds of thousands—I mean, is it soon to be millions of women—who have had their images used, you know, commandeered, we'd say in a legal sense, and turned into this sexually explicit material? How——

Mr. COLMAN. Right.

Senator HAWLEY [continuing]. Are we going to empower them?

Mr. COLMAN. You know, it's quite simple. You mentioned CSAM imagery. There's a really nice framework in both national and also State-level regulations in this space. When you upload something on, for example, YouTube, it's checking for a few things. It's checking for violence. It's checking for underage imagery. It's checking for, are you uploading the latest Drake song? That's because of regulations.

So, to scan for generative media would just be another check within that same flow. It's nothing new, nothing novel. It just needs your teams to actually push it forward to require the platforms to protect the consumers, because in the absence of this, we have things like community notes, which only actually cause anything once things have been shared 100 million times. Or, worse than that, we have content moderation teams, which we've seen be slashed, and they don't really do anything at all. So, the challenge here right now is that the technology exists; we have folks on this dais who can actually solve for it. We need regulations to require the platforms themselves to use us, the same way they're required to scan for underage imagery.

Mr. GUPTA. Sorry. Really quick, a point on——

Senator HAWLEY. Sure.

Mr. GUPTA [continuing]. That. I think it's important to understand that, for, like, a deep nude image—they're shared on, like, WhatsApp and things like that—those are end-to-end encrypted. You can't detect that. It doesn't make any sense. To solve that specific deep nude, you need AI poisoning, which is part of the defense in depth. Anytime you upload an image to Instagram, you can poison it so that, if someone tries to deep nude it, it turns out as garbage. So, specifically for deep nudes of, like, images posted to Instagram, AI poisoning is that solution.

Senator HAWLEY. Yes. You know, I hear what you're saying, and all of that sounds good, and I hear what you're saying about the platforms' current obligations and their current rules that they have in place, for instance, to detect CSAM and so forth. But the problem is that this Committee and other Committees have heard mountainous evidence that these same platforms are absolutely awash in CSAM that is not digitally created. It's not synthetic. It's, you know, quote, unquote, I mean, "real." It's actual people, which is even worse.

And, I mean, they're just—so, they say that they're trying to do their best, but it is absolutely—the internet, particularly Facebook, Instagram—absolutely overrun, which brings me to what I think has got to be part of this conversation—is that we have got to allow Americans, ordinary Americans, individual citizens—we have got to allow them to get into court and to hold accountable, legally, companies who are producing or hosting this content. If we don't do that, I don't see how we change the incentive structure. If Instagram fears that it's going to get a billion-dollar jury verdict against it, they'll adopt all kinds of new technology. But if they don't, they won't. Thank you, Mr. Chair.

Chair BLUMENTHAL. Thanks, Senator Hawley. Senator Klobuchar.

Senator KLOBUCHAR. So true, Senator Hawley. Right now I'm going to focus on elections, but I will say those were startling numbers, Mr. Colman, and I think it is just what we are seeing, both real people and AI-created people. It's one of the reasons that we got the SHIELD Act through here, which is not the liability issue, that I also support, that Senator Hawley was mentioning, but also getting the information to law enforcement and the like, to be able to make it easier to go after these perpetrators.

And we can just sit here and do nothing; we can pass resolutions; we can—but unless we empower people to go after these cases and then equally make liability, it's just going to get worse and worse, and at some point the public will have had it, and I don't know if that's what—this year, but it's going to happen. And so I'm telling my colleagues this.

So, let's go to a few things here. The bill that I mentioned, that Senator Hawley and Coons and Collins and Bennet, Ricketts, other, and I have—could you tell me, Mr. Colman, how AI has this potential to turbocharge election-related disinformation and why we can't just rely on the disclaimers and watermarks? I think you can do that for a set of it. I don't think you should do it for all uses of AI, and we have a labeling bill that I think differentiates that. But for this really bad stuff that Secretary of State Scanlan was referring to, tell us why it's not enough—this is softball, but—to run a whole thing and have a little label at the underneath, when they think it's the actual candidate but it's not. So——

Mr. COLMAN. We agree on that. I think that, to paint a larger picture, what we saw during the primaries was a single static deepfake, prerecorded, kind of a one-to-many attack. It didn't change. It wasn't even live. Imagine a world where that was a one-to-one attack, where instead of it being prerecorded, it was actually live, and instead of being from one to many, it was one to one, where it's coming from your husband, your wife, your boss, saying, hey, Ben, we need you in the office at six a.m. I know it's a voting day—or to an election official, hey, we're moving your precinct. We need you to be across town, 3 hours away.

And so that's where this is going to go. It's not going to be a single prerecorded, you know, arguably medium-level deepfake. It's going to be a real-time, custom deepfake in conversational language, having people do all kinds of things at all levels of the election system. So, on our side, we see this as a massive issue, not just in the U.S. but globally. And what's great here is on the dais

we have different technologies all solving very much the same issue. It's all possible now.

We have large companies, we work with large banks, large government groups, large media organizations that are thinking long-term and already solving for this. We have banks scanning incoming phone calls, every single phone call. We don't have anyone protecting average consumers: my parents, my grandparents. They just don't stand a chance. With other technologies——

Senator KLOBUCHAR. Right.

Mr. COLMAN [continuing]. Whether it's CSAM or, for example, a computer virus—they don't have to be experts. They don't have to tell—ransomware or NAPT. They just know that their email provider will actually block it for them. We're looking for the bare minimum there, which is just letting us know that maybe something might be fake and then allowing us to decide maybe we don't want to see it, maybe it won't go viral. But right now, the things that are the most extreme go the most viral. The platforms that do think about this are already solving that, using technology like ours.

Senator KLOBUCHAR. Right. Very good. And I do want to note that drafting this bill, the deepfake bill, wasn't easy. We had to look at allowing satire, right, and all these kind of things, within the framework of the Constitution, and having Democratic and Republican lawyers look at this, to figure out what gave us a chance. I just think if the platforms can point to something, as opposed to laws that aren't quite on point—which 11 States have done for States but not for Federal—and say, we've got to take this down, we're going to be in a much better place than we are with a little label that they may not even notice. And it also—the labels—you know, I think it's important for some of this, but I don't think it can be the only answer.

Mr. Scanlan, you know, birthplace of democracy—no, kidding—spent a little time in your State, there. I know you cherish democracy very much. Could you talk about what other Federal support would be helpful in taking this on, in addition to stronger laws?

Mr. SCANLAN. Secretaries of State for a good decade now have been dealing with misinformation and disinformation, generally, and that takes on many different forms. And there's no question that today, voters receive their news in different formats than they did 20 years ago. And a lot of that news is electronic. It's on their cell phone. Many voters believe exactly what they see on that format, on that media, without question. So, in addition to whatever might be appropriate to help States recognize and put brakes on malicious technology, in terms of deepfakes, I think we have to spend a real strong effort on the fundamentals of transparency and helping voters and educating voters on the election systems and how they run and what the checks and balances are that are protecting them in the polling place.

Senator KLOBUCHAR. Right. I've always found it interesting, like, those Baltic States on the border with Russia. They were putting up misinformation, lying things, and they, over time—because of education, they kind of have seen through some of it. It is possible. It can't be our only answer, because of what everyone's being ex-

posed to, but I think it's a good point, and we have the Election Assistance, of course, Commission.

But I did want to say I appreciate, as a Republican Secretary of State, how seriously you and the attorney general and others in New Hampshire took this egregious breach with the guy that did an interview afterward. Maybe they should've hired a mime instead of a magician, but in the end, I just think that we've got to make clear there's consequences when this happens, as well.

I have other questions—I don't want to go over my colleagues' time; I already have, about—that I'll ask you on the record, Mr. Ahmed and Mr. Gupta. Thank you so much for being witnesses today, but I just—we have to be as sophisticated as the people that are messing around with our democracy and our laws, and that's why we've got to get these bills done. Thank you.

Chair BLUMENTHAL. Thanks, Senator Klobuchar. Senator Hirono.

Senator HIRONO. Thank you, Mr. Chairman. Maybe this is something that Secretary Scanlan can talk about. We have laws that we're contemplating passing in this Committee, as well as in the larger Committee, the full Committee. But where does educating the public come into play, to let them know that, as we are approaching or we're already in an election situation, that they should expect to experience AI-created audio, video fake stuff, that—where does educating the public to the fact that they will be subjected to all of this come into play, to inoculate them against the impact of this kind of fake, deepfake material?

Mr. SCANLAN. I think the States are best suited to deliver the message——

Senator HIRONO. Are they, though?

Mr. SCANLAN [continuing]. To their voters. I believe they are. Now, I believe that there is a role for the Federal Government to assist them in that, because elections are run differently in every State.

Senator HIRONO. Yes.

Mr. SCANLAN. But there is a—I mean, this issue with AI and the impact that it can have—it could actually bring down an election, if it's done successfully, and that's a national problem. It's not a State problem. You know, the States, I think, are prepared to help deal with it, but the narrative has to be uniform.

Senator HIRONO. Well, what I'm getting at is that, you know, people in—the use of AI, I think, is going to be very prevalent in this election, upcoming election season. The voters should know that they will be subjected—they may not know it. They may not even believe that it's happening to them. Something happens, and then, after the fact, you have a press conference and you say, oh, there was a fake President Biden telling people not to vote. That's after the fact. How do you inform people that they should be aware? Are States doing this, and does that play a role? That's the question that I have, because I don't know that that is happening in the States. It's usually after the fact that they are informed.

Mr. SCANLAN. In New Hampshire, we're trying to raise the level of awareness of voters so that they know what to expect during an election cycle. I don't know that we can do any more than that. Some of these deepfakes can be incredibly real, and, you know, I don't know how we—you know, how we deal with that in real time.

Senator HIRONO. Mr. Colman, do you have something to add?

Mr. COLMAN. Yes. You know, at Reality Defender, we don't sell directly to consumers. We sell to large companies: for example, large investment banks. And large investment banks have an internal challenge educating their employees, whether it's deepfakes or ransomware or spam or different kinds of scams. And what we've seen work—and the only thing that works—is to actually, you know, try and scam the employees; obviously, teach them what's happening and let them look back and see what happened.

An example with phishing email campaigns—one of the most standard tools to educate employees about phishing campaigns is to actually send them phishing campaigns and then afterwards ask them if that was real or fake, and do it over a continuous basis, over weeks and months. And I would imagine—you know, I can't speak to whether it's State or Federal; that's your world, not mine—but I can imagine, given that we're all talking about cybersecurity education and hygiene, a very similar approach could work on a very large level.

Mr. GUPTA. I do think it's worth pointing out that deepfakes—image, audio, video—this is a new paradigm, right. It's a new reality. There is evidence from the University of Maryland School of Public Policy that, when voters are informed about what policy looks like, there is wide bipartisan support for Federal regulation. There just is, right. I believe it's 89 percent support, across the board. And I can share that report with you, if you're interested.

Senator HIRONO. I think one of you said that there are countries—Taiwan, even China—they already have legislation to protect against use of AI-created deepfakes. So, are any of them applicable, do you think, to our country? Mr. Colman.

Mr. COLMAN. Absolutely. You know, and just at a high level, you know, the majority of use cases for AI and generative media are great. They're going to help the world do a lot of great things. They're going to help create medicine faster.

Senator HIRONO. Yes.

Mr. COLMAN. Solve all kinds of societal issues. And this is one very small issue that has very large, asymmetric penalties, as they say. And so what we're seeing in other countries, whether it's in Taiwan, Singapore, or Japan—even China, but also UK, European Commission, Canada—is the first step is—the bare minimum is just indicating that something may be fake. Not saying they're blocking it or flagging it or damaging the user, but just saying it may be. And, you know, I—full disclosure. I'm a Google alum. But Google is taking an interesting approach of, one, requiring uploaders to confirm whether or not they're generative media, and if you confirm it and it's later found out, you might lose your account.

And so certain platforms, certain organizations, are thinking long-term about this and saying, it's going to happen anyways. It's already working in other parts of the world; it might work here, as well. It is a stepwise approach, along with education, which I think you were mentioning with the secretary of State, as well. But I think there are certain stepwise approaches that are absolutely applicable here, in a year where elections are going to be paramount.

Senator HIRONO. I think that these platforms will start paying a lot more attention to the content on their platforms if we start to move toward eliminating Section 230 liability protections. Thank you, Mr. Chairman.

Chair BLUMENTHAL. Thanks, Senator Hirono. Senator Whitehouse.

Senator WHITEHOUSE. Thanks, Chairman. I just wanted to find out from each one of you if you think that there is a particularly good source—like, for instance, New Hampshire's Attorney General—or a particularly good article or analysis on where gaps are in the criminal law that should be plugged, in order to deal with the problem of deliberate and malicious AI fakery. Do fraud laws need to be adjusted? Does it need to become a RICO predicate? What are the—and I don't know—go ahead, Mr. Gupta, but also I'd like to hear from any of you, if you're not the expert but you have somebody you know or admire or think does good work in this space, if you could let us know, because we're trying to——

Mr. GUPTA. Yes. Definitely. Again, this report from the University of Maryland School of Public Policy, by Stephen Kull—it's really in depth, and it talks about what the people think should be done. Right. They actually take time to educate the people about certain policies——

Senator WHITEHOUSE. Yes.

Mr. GUPTA [continuing]. And then poll them about: should there be an independent Federal organization? What kind of laws should exist? How extreme should those laws be?

Senator WHITEHOUSE. Yes.

Mr. GUPTA. I am not the expert. Legislators are the expert. But I fundamentally believe that legislators should be informed by the public's opinion, and so I——

Senator WHITEHOUSE [continuing]. Report would——

Mr. GUPTA [continuing]. Would highly recommend that.

Senator WHITEHOUSE [continuing]. Be one good place to look.

Mr. GUPTA. Yes.

Senator WHITEHOUSE. Mr. Secretary.

Mr. SCANLAN. Thank you, Senator. So, I believe New Hampshire is probably the first State, relative to elections, where the attorney general is investigating and hopefully prosecuting the individuals responsible for——

Senator WHITEHOUSE. Yes.

Mr. SCANLAN [continuing]. The deepfake, and I think that it's probably going to take that exercise and that experience to figure out where the gaps are.

Senator WHITEHOUSE. Yes.

Mr. SCANLAN. And I'd be happy to report that to you when they complete that.

Senator WHITEHOUSE. That would be great, and if you don't mind, when you get home to your Granite State, let them know that we were asking about the attorney general and if they could let us know or have their policy person check in, that would be helpful.

Senator WHITEHOUSE. Mr. Colman.

Mr. COLMAN. On the topic that Ranking Member Hawley mentioned, around deepfake nonpermissioned or nonconsensual pornog-

raphy, what we've seen time and time again is that students, young men, are creating deepfake imagery of women in their classes. And what's double challenging here is that while it might break rules within the schools, it's not breaking any local laws.

Senator WHITEHOUSE. Yes.

Mr. COLMAN. And so the challenge of this potential patchwork of laws, which doesn't exist in most States, where they're committing—initially they'll get suspended from their school; they won't be arrested. So, this is an area that can follow other types of emerging regulations and also the penalties around CSAM imagery, because I would argue that any image that's of an underage person that's nude definitely is effectively CSAM, whether it's real or fake.

Mr. AHMED. I could perhaps just add a couple—I think these are all great points. I think we could probably look back to 1999, 1998, when the internet was really young. We had, you know—piracy followed a very similar path. If you try uploading an NFL video on YouTube, it won't make it past the upload screen today. That is because there is watermarking technology that is broadcasted. So, there's also very strict penalties. The piracy law makes it very strict in terms of—if you proliferate, if you upload pirated data, there are consequences. There are letters sent to your home.

So, I think, like, we've solved some of these issues in the past already. You know, the folks who are doing, like, the trust and safety work at YouTube have probably been dealing with a lot of this stuff—non-AI-generated, but still, like, you know, flipped images, mirrored images, et cetera; altered forms of the same material being uploaded. So, I think this is—we could take a lot of inspiration from there.

Mr. GUPTA. One more thing, Senator. I think it's really important that the United States military, the intelligence communities, are funded properly to help integrate this technology, because if the U.S. Government doesn't know what's real and what's fake, then we have a really big problem.

Senator WHITEHOUSE. Thank you very much. Thanks, Chairman.

Chair BLUMENTHAL. Thanks, Senator Whitehouse. Thank you. I'm going to ask some more——

Senator WHITEHOUSE. And thanks for——

Chair BLUMENTHAL [continuing]. Questions.

Senator WHITEHOUSE [continuing]. Being patient with our schedule today. We had a——

Chair BLUMENTHAL. I explained that.

Senator WHITEHOUSE [continuing]. Group of visitors from the House who had something to deliver.

Chair BLUMENTHAL. Very well said. I have a few more questions, and I apologize for the lateness of the hour, so I'm going to try to be quick in my questions. Mr. Ahmed, I think you had a comment on my question. If you can't remember it, you're more than forgiven, and I want to prompt you with another question along the same lines.

We were discussing watermarking. The point was made that this labeling, watermarking, was insufficient in and of itself. The idea of an independent entity would be not necessarily only to set a licensing regime but also establish something more than just simplistic watermarking, as all of you have suggested. You actually

use voice cloning software. The voice cloning software, incidentally, from New Hampshire, I understand, was from a company known as ElevenLabs, or it was created using software from ElevenLabs.

Most of these voice cloning tools don't require the consent of the person being impersonated. You suggested earlier, I think, the idea of a public data base, traceability. Could you expand on that point? Because it seems to me we've been talking about essentially defensive measures of a very simplistic kind. If we can use the technology to flip the model, so to speak—there is technology that can be used to apprehend and trace the bad guys—that might be a strong deterrent.

Mr. AHMED. Yes, absolutely. That's exactly how we're thinking about watermarking. You know, and to the earlier point, like, the watermarking that we're talking about here is imperceptible. It's inaudible. It's deeply sophisticated. It's a neural network that's embedding watermark, so it's very difficult to replace or remove these watermarks.

I think, you know, we—this kind of indirectly kind of answers a bunch of questions and topics today. I think we need to hold a lot of generative companies accountable. You know, it shouldn't be the case where someone can go in and—as unsophisticated as these attacks are at the moment, they're largely there because they're not even writing code. They're actually going to online websites, sometimes even to the Apple's app store and downloading an app and effectively doing something there. So, I think, you know, we could start at, like, the generative models, modeling perspective, there.

I think the idea of traceability is extremely important. You know, the world that we don't want to live in is a world where everyone just shuts down, creates their data lake and says, this is our data. No one else can touch it. No one else has access to this data. Right? And you're already seeing that with Reddit. You're seeing that with Twitter, where the API access is being shut down. You have other, you know, companies that rely on that data, to suffer.

So, I think the twofold answer to your question: one, the generative models, if we can create some sort of a watermark—and there's tons of research in this area—where a subset of the data, even if it's watermarked, the watermark persists through model training; shows up on the other side. You can see, like, oh, this was created with this source of audio, right. And I think that's extremely important.

The second is this idea of this public data base. I think a lot of this does come to education. You know, when I was in school, you were told how to use the internet. You were told how to chat online. You were told where to go. I think the world has, like, significantly changed. I think people—like kids, adults, et cetera; people who work in enterprises—need a place and a source where they can look at vulnerabilities.

And I think, you know, as you grow up—you know, I studied computer science. In our forensics class, we would look at, like, reports like Google, Apple would publish as, like, incident reports and try to analyze them. I think a public data base of all incidents, that is very easy to find, neatly categorized, serves as good educational material trying to demonstrate how attacks were created, and that basically can be a great resource for education but also

a great resource for understanding, you know, where the gaps are in terms of holding these generative companies accountable and making sure that they're not that easy to access.

Mr. GUPTA. I would like to just commend that statement and also let you know that that currently exists with the DARPA AI Force. Like, the DARPA is working on that exact idea, and I think it is a great idea.

Chair BLUMENTHAL. Thank you. Mr. Colman.

Mr. COLMAN. I think everything they're describing is a fantastic start but presupposes that, again, everyone's going to follow the same rules. You know, whether it's downloading open-source software or a State-level actor using software that does not follow the rules or has hacked the rules—best case is, you have a watermark or a cryptographic hash that's wrong, that gives a false sense of security; and worst case, you don't have anything at all, because the bad actor doesn't care about the rules and doesn't follow them. And so our focus is on the probabilistic view that doesn't need any watermarks. And I think this is where we could all work together, but I just want to share that there's two sides of the same coin.

Chair BLUMENTHAL. Yes. I'm assuming that no one follows the rules, voluntarily.

Mr. GUPTA. I do think, really, it's more about that Swiss cheese model, if you're familiar with that, right, that you have all of these different things in place, and stuff's going to fall through the holes. And deepfake detection companies like ours—we're good at deepfake detection, right. We have really high quality accuracy. We present heat maps and information and probability scores, and all of that's delivered to Big Tech companies at scale; militaries at scale. Like, that's what we do, right. But that's not enough—by itself is not enough. You need everything.

Chair BLUMENTHAL. You need some kind of punishment when people fail to follow the rules.

Mr. GUPTA. That, too. And you need to know what's real and fake, right? This is counterfeit truth, and like we have counterfeit currency, right——

Chair BLUMENTHAL. Well, you need to be able to know and to prove——

Mr. GUPTA. And punish.

Mr. COLMAN. Mr. Chairman——

Chair BLUMENTHAL [continuing]. To prove, just as you would in counterfeiting, just as you would somebody speeding. We don't assume that everyone's going to follow the speed limits.

Mr. GUPTA. You need police with radar detectors.

Chair BLUMENTHAL. Enforceability.

Mr. GUPTA. Yes. Right.

Mr. COLMAN. But——

Chair BLUMENTHAL. Yes.

Mr. COLMAN. But I think, beyond all this—these are all fantastic ideas that, at some level, we'll solve for. But I think we can all agree we need to start somewhere. We need to start somewhere with baby steps. We can start adding additional things, but we still haven't started yet, and we're months away from an election year.

Chair BLUMENTHAL. Well, you have just crystalized or expressed the anxiety that many of us feel, because we are approaching the

election. As I mentioned at the very outset, we're facing a deluge of this stuff. And, by the way, it's not the first time that we've faced distorted electioneering or fake ads. When I first started out, our great fear was on the Sunday before election someone would go around with mimeographed pieces of paper and put them on the windshields of cars parked at church, without identifying who it was from, but distorted—images of the candidate doing something terrible.

So, the idea—and, you know, you go back to the founders. They were worried about false electioneering, as well. As secretary of State in New Hampshire, you have a concern with making sure that elections are fair and honest. This problem didn't just arise. But you're absolutely right. We're facing an election where we need to take some steps right away. I'm not going to say they're baby steps, but steps right away.

Mr. GUPTA. I would like to caution us against taking—like, adopting the Chinese model approach. There's a really great book by a Columbia law professor, Dr. Bradford, where she outlines the Chinese State-driven approach, the European rights-driven approach, and the historical United States market-driven approach. So, something needs to be done, but a state-driven approach has serious and significant harms to the public, and I want to caution us against adopting regulation that China has put in place.

Chair BLUMENTHAL. Let me kind of come back to one of the key questions, and I mentioned earlier that my belief—Senator Hawley agrees—that Section 230 does not apply to AI, and therefore we have a legal basis to hold Big Tech accountable here. To what extent does Big Tech know or should it know that these kinds of con artists are using their platforms?

Mr. GUPTA. I can say that we are currently engaged with the Big Tech companies. The Big Tech companies are deploying our technology to fight against deepfake misinformation already. That's already happening.

Chair BLUMENTHAL. Well, they're not doing a very good job of it.

Mr. GUPTA. Yes. I know. They need to use it more.

Mr. COLMAN. Mr. Chairman, I'd say some large tech companies are thinking long term about this. Others of them, without naming any specifics, have completely decimated their teams that are focused on trust and safety and have cut their budgets on actually using software from any of us. A lot of this is public. It's all in, you know, the—yes.

Chair BLUMENTHAL. I'm guessing, from what I've ready publicly, that one of the companies that have cut their staffs is Meta.

Mr. COLMAN. I won't comment on that, but I will say that every one of them needs to expand these teams. They need to see them as an expected requirement from government and not a cost function, because we've seen the most recent layoffs. These are the first teams to go. And they're not just 10 or 20 percent of the teams. We're talking about the whole teams: everybody in trust and safety; everybody at identity KYC; everybody in fraud—completely gutted.

Mr. AHMED. You know, there are many, many great startups. You know, I feel really small sometimes. And I share the frustration, because we're building the technology, and, you know, for a

while, we, you know, tried to get it to the platforms, and we realized the elections are, like, you know, months, a year—years, when we started; months, now, at this point—and this is one of the reasons why we kind of skipped the line. We decided, Okay, you know what? Like, the consumers—they need a tool that they can access today. We need to give them a way to get to the tool themselves.

This is exactly—where we are right now is 1998. GeoCities was—like, you know, people are building websites in GeoCities, and they're shipping them, and there's malware, there's spyware, there's all sorts of stuff. We're now at the point where we can expect a browser to pop up a red screen. That happened in the mid–2010's, right? Like, that's very recent news. But for a long time, we lived on the internet, and it was still very early days.

And so our goal and, you know, what we've shifted our focus toward is providing a tool that anyone could access, not just for enterprise, not just for, you know, the companies that are out there, but for normal consumers, so they can go and they could validate against, like, YouTube videos, TikTok, Twitter, Tiklike, et cetera. They can go all over the place and try to validate that. And our goal is trying to get this technology, ourselves, as much as we can do, to the consumer, as quickly as we can. Yes.

Mr. GUPTA. One——

Chair BLUMENTHAL. I am taking from these answers that the social media companies know, or should know, because what you have said, basically, whether it's cutting their teams or using your software or inventing technology that can trace the deepfakes—that they have the capacity. They know or they should know when their platforms are being misused. Mr. Gupta.

Mr. GUPTA. One more quick thing, here. You know, I want to highlight that issue of scale, right? The firing of the trust and safety teams is a tragedy. It is. But the amount of deepfakes and the amount of disinformation is not going to be solved by hiring those teams back. They would have to hire 10, 20 times more people. AI must be fought with AI. And for every deepfake you see on the platform, there are hundreds more that were removed and not submitted because they were filtered out.

Chair BLUMENTHAL. Thank you. You may recall the devastating wildfires that spread across Hawaii. As I mentioned earlier, the Chinese Communist Party decided to spread the disinformation that the disaster was the result of a United States weather weapon test, as it was called, weather weapon. This conspiracy theory showed Beijing's willingness to directly meddle in American affairs. I'm sure there have been others, but it was supported by AI-generated images. Maybe you can talk about some of the dangers to the United States from our foreign adversaries, not just within the country like the New Hampshire primary person.

Mr. COLMAN. Yes. I'll give a very vivid example. A few months ago, there was shown online, on X, on Twitter, what looked like an explosion of the Pentagon. And, you know, part of the reason—because there're no regulations to automatically scan for it on upload, it took millions of shares and reshares for it to be flagged by Community Notes. By that time, it led to a $100 billion flash crash in the market. Now, the market did come back, but this is a really simple image which arguably was a diffusion-based deepfake. We

detected it doing what's called frequency domain analysis, but this is an example of how you cannot only move an election but also move markets with a single photo, correctly placed on the right social media platform, and then just letting it go viral on its own.

Mr. GUPTA. I would like to highlight the difference between misinformation through a telecom, like the Biden robocall, and misinformation through the platforms. Those are very different things. They are operated very differently. I believe Reality Defender works very closely with financial services and telecoms to solve that problem. Deep Media works very closely with the platforms to solve that problem. We also work very closely with the Air Force Cyber Command Division, 16th Air Force Unit, as well as NASIC, as well as the U.S. Army, to fight foreign interference from a military standpoint. So, while in the interest of national security I would like to not go into specifics, we are monitoring for Chinese and other near peer adversary deepfake-based misinformation and narrative redirection campaigns.

Chair BLUMENTHAL. Let me conclude my questions by going to Mr. Scanlan. You mentioned the possibility that AI could be used to "bring down an election." Maybe you could expand on that point.

Mr. SCANLAN. What's most important in the election process is that the voters have confidence in the outcome and the results of an election. And as long as they are confident, they are going to participate. And New Hampshire and Minnesota consistently are among States where we have very high voter participation. But if voters start believing that the government is corrupt or the election outcomes are not accurate, they've been manipulated somehow, then participation is going to decline. And it only takes one really serious event where an election at least has the appearance of having a major breakdown in terms of the outcome to throw doubt in the voting population. And I think that's a really, really significant concern.

You know, that is probably the most fundamental, important thing that I perceive in my role as secretary of State—is to make sure that an election is not messed up; that, you know, the voters believe and know that it was run fairly and accurately, to the highest standards possible. And if we lose that, it will be very, very hard to get it back.

Chair BLUMENTHAL. Thank you. I'll open it to any final comments that anyone may have, if you haven't had an opportunity to say something.

Mr. AHMED. No, I think—you know, I'll echo the point confidence is—voter confidence is so key. You know, again, we have a list of these attacks that occur other places, in politics, et cetera. One to point out is actually in—the British opposition leader, Sir Keir Starmer, was—you know, there's a deepfake of him berating his staff, right? That gives an impression to voters that is not correct. You know, it's not fair to him. And, you know, overall, like, once—like, you know, what my colleague said earlier—when everything in the world is fake, you don't know what's real anymore, and I think that's extremely important.

Mr. COLMAN. Yes. I'll just add and reiterate that I don't think any single solution is ever 100 percent, and the opportunity that exists with developing the world's best technology can also apply

here, as well. Really excited for the collaboration among startups and big companies but also our elected officials in solving this very important issue.

Chair BLUMENTHAL. Thank you.

Mr. GUPTA. I would like to highlight two things. I definitely agree with Ben that no single solution works. I believe the defense-in-depth solution provides a comprehensive way forward. I also want to highlight that, again, there are frameworks for legislation. There's the State-driven approach with China, the rights-driven approach with the European Union, and the market-driven approach with the United States. A market-driven approach is good for the generative AI companies and the platforms, and it's good for the U.S. people. It is not dissimilar to the proposed legislation that you've put in place. I actually think that solves a lot of these problems. But to me, it's about internalizing these negative externalities so the AI ecosystem can grow safely, largely, quickly, and we can all get rich.

Chair BLUMENTHAL. Mr. Scanlan, Secretary of State, your synopsis or your summary of the dangers here I thought was very eloquent and powerful, but I know you're on the firing line, literally every election. Every time people go to the polls, that issue of trust and credibility is there, and it relates not only to those of us who are candidates but anybody who goes to the polls and wants to have confidence that the outcome's going to have integrity. Thank you for being there.

Thank you all. This was an excellent and very informative and helpful session. Again, my apologies for the lateness, and this hearing is adjourned. The record will stay open for a week, in case my colleagues have additional questions for you in writing.

Chair BLUMENTHAL. And the hearing is adjourned.

[Whereupon, at 5:09 p.m., the hearing was adjourned.]

[Additional material submitted for the record follows.]

APPENDIX

Submitted by Chair Blumenthal:

Written Testimony of Zohaib Ahmed, CEO and Founder of Resemble AI

For a hearing on "Oversight of AI: Election Deepfakes" Before the Judiciary Committee

Subcommittee on Privacy, Technology, and the Law
United States Senate
April 16th, 2024

Introduction

Chairman Blumenthal, Ranking Member Hawley, and Members of the Committee, thank you for the opportunity to discuss the oversight of AI as it relates to understanding the impact this technology can have on elections with you. I'm Zohaib Ahmed, CEO and Founder of Resemble AI.

Resemble AI is a research and development lab focused on the creation of Generative Voice AI models and is being used by some of the largest media, gaming, financial and telecom companies as well as content creators.

Our team has spent the last five years developing and researching AI voice technology and we are uniquely positioned to understand both the remarkable potential and possible risks associated with the rapid advancement of voice synthesis and cloning capabilities. It is with this balanced perspective that Resemble AI has created innovative solutions to address the emerging challenges posed by unauthorized or unethical uses of voice cloning technology.

Given our tenure in responsible voice cloning technology, we started building speaker identification, watermarking, and deepfake technology to enable safer deployment of voice AI products. Over the last nine months, we have opened up the research on speaker identification, watermarking and deepfake detection to a broader audience to help ensure that Generative AI is safely and responsibly deployed around the world.

I'd like to address the critical role of responsible technology practices in shaping the future of AI. We have always maintained a commitment to ethical standards, including transparency, privacy, and security; and these principles guide our product development and user engagement.

In this written testimony, I will share some of the technologies that we have developed since Resemble AI was founded in 2019, especially around watermarking and deepfake detection, and my recommendations around 1) transparency and disclosure, 2) safeguards and mitigation, and 3) integrity verification.

Ethics and Consent

At Resemble AI, we've always held ourselves to an exceptional standard of ethics. To uphold our mission to help the private and public sector use generative AI for good, we developed the following guardrails early on to maintain our ethical standards, including:

- a dedicated team that is responsible for ensuring that our voice cloning technology is used ethically and responsibly
- policies and guidelines that we require our customers to adhere to, including a consent line that must be recorded in the same voice that is being used to generate new audio
- policies and guidelines are designed to ensure that our technology is used in a way that is respectful of people's privacy and data rights and that it is not used in a way that might cause harm
- technological improvements that detect AI-generated media including our neural watermarker and AI-speech detector

At Resemble AI, we require users to provide a recording of a consent clip in the voice they are attempting to clone. If the voice in this clip does not match the other clips, the user is blocked from creating the AI voice.

When recording, Resemble AI requires the user to say an array of particular sentences in your own voice. Misuse of this can be easily detected by our algorithm. Once the voice is created, the user owns all rights to that voice. We do not use that voice data to train other models, nor do we resell the voice data to third party companies.

For customized solutions, we work with companies through a rigorous process to make sure that the voice they are cloning is usable by them and have the proper consents in place with voice actors.

Materials used through integration of Resemble and related metadata must be produced by the publisher itself, correctly licensed from the third–party rights holder, used as allowed by the rights holder, or legally used in any other way.

In our terms of service, we state that you can not use AI voices built by Resemble for:

- claiming to be any person, company, administration, or entity without explicit authorization to make this statement and/or impersonating to gain illegal information or privileges;
- propagating hate speech;
- discrimination, libel, terrorism, or violent activities;
- spreading unattributed content or misrepresenting sources.
- exploiting or manipulating children;
- making unsolicited phone calls, vast communications, postings, or messages;
- deceiving or deliberately misleading people;

Watermarking

After creating open source speaker identification with Resemblyzer several years ago, in 2023 we took it several steps further by introducing the Neural Speech Watermarker, an "invisible watermark" that tackles the malicious use of AI generated voices. With a deep neural network watermarker, the data is embedded in an imperceptible and difficult-to-detect way, acting as an "invisible watermark."

To deploy safe neural speech in the wild, Resemble AI introduced the PerTh Watermarker, a deep neural network watermarker. It uses machine learning models to both embed packets of data into the speech content that we generate, and recover said data at a later point. Because the data is imperceptible, while being tightly coupled to the speech information, it is both difficult to remove, and provides a way to verify if a given clip was generated by Resemble. Importantly, this “watermarking” technique is also tolerant of various audio manipulations like speeding up, slowing down, converting to compressed formats like MP3, etc.

Our model is called PerTh, which is a combination of Perceptual and Threshold. It was designed around the concept of exploiting the way we perceive audio to find sounds that are inaudible, and then encoding data into these regions. Further care is taken to ensure we can extract the embedded data from any part of the audio (aside from silence), and that the data is encoded into frequencies most common for speech. This ensures our data payload is difficult to corrupt with common audio manipulations.

We use Psychoacoustics (the study of human sound perception) to find the best way to encode the data. One psychoacoustic phenomenon is that we have varying sensitivity to different frequencies — this means we can embed more information into frequencies we are less sensitive to. Another, more interesting effect called “auditory masking” also exists, in which quiet sounds nearby in frequency and time to a louder sound are not perceived. As a result, the masking sound (the louder sound) produces a “blanket” in amplitude-frequency-time space that drowns out other sounds beneath it.

Further research of our watermark supports the traceability of data. Our watermark has proven that it can persist through model training, therefore the output of a Generative AI model that uses watermarked data can be traced back to the source.

Deepfake Detection

Resemble AI's Generative Product, encompassing Voice Cloning combined with text-to-speech or speech-to-speech technologies, is extensively utilized by enterprise customers for conducting red-team exercises. These exercises are critical for security testing, where the technology's ability to replicate voices accurately is used in simulated attack scenarios to assess and improve organizational security measures. By creating realistic voice interactions, the technology aids in identifying potential vulnerabilities, thereby enabling companies to enhance their security protocols and safeguard against actual voice-based security threats.

In July 2023, we launched Resemble Detect, our advanced deepfake detection tool, which utilizes a cutting-edge neural network to differentiate between real and fake audio files, addressing the increasing concern of deepfakes in various sectors. Resemble Detect is designed to work in real-time, providing up to 98% accuracy in exposing deepfake audio, thereby safeguarding the authenticity of voice content.

For telecom providers, Resemble Detect can uncover social engineering fraud attempts using cloned voices in phone calls. These can be flagged for enhanced screening to protect customers against pretexting scams.

To further support our enterprise customers, many of whom serve as a gateway to consumer safety, we launched our Deepfake Detection Dashboard to help security teams monitor deepfake activity in a central place.

We acknowledge that consumer education and awareness is another critical piece to addressing the risk of misinformation. Starting last fall, we also began publishing weekly Deepfake Incident Reports to bring transparency to the use of deepfakes using our Detect technology. We make these available on our blog and offer our reporting to journalists to accelerate the identification of deepfakes in mainstream media channels.

We also joined the FTC's Voice Cloning Challenge and signed Canada's voluntary AI code of conduct, underscoring our commitment to accountability, safety, fairness, transparency, human oversight, and robustness in AI systems.

Our free, real-time Deepfake Detector tool is available for anyone to combat fraud and promote the responsible use of generative voice technologies. With this tool, you can quickly verify the authenticity of widely circulated audio content, making it a valuable asset for journalists, content creators, and the general public who are often on the frontline of combating misinformation.

Our solution also helps enforce the recent FCC ruling against robocalls by providing a means to verify the authenticity of audio content. It can serve as a line of defense for businesses and individuals who wish to ensure that the voices they are hearing in calls are genuine and not AI-generated deepfakes intended to deceive.

And just last week, we added real-time deepfake detection for meetings, starting with Google Meet, that leverages our Detect technology. By integrating into Google Meet, our Detect solution is now available on a platform that people are already familiar with and it provides them with a seamless detection experience that can give them a piece of mind they are protected from deepfakes and ease AI fraud concerns.

Policy Recommendations
I believe your legislative framework is a comprehensive approach and positive step towards effectively regulating AI. Given the rapid advancements in generative models and the surge in synthetic voice technology, especially during this crucial election period, I would like to propose additional recommendations for your thoughtful consideration as you continue to refine legislation in this area:

1. **Transparency and Disclosure**

First and foremost, we support the proposed legislation that requires clear labeling on AI-generated content in the election process. Similarly to how disclaimers appear at the end of

political ads, consumers should be made aware that they are interacting with an AI model or AI-generated content.

To take it one step further, we propose the creation of a public database where all AI-generated election content is registered, allowing voters to easily access information about the origin and nature of the content they encounter.

Furthermore, we believe that voter education initiatives are crucial in promoting transparency. Resemble AI recommends the development of public awareness campaigns that inform voters about the existence and potential impact of AI-generated content in elections. These campaigns should provide clear examples of how AI can be used in election-related communications, both positively and negatively, and equip voters with the tools to critically evaluate the information they receive. Language translation is an example of how voice cloning technology can be used responsibly in an election year.

2. Safeguards and Mitigation

To adequately safeguard against misinformation, particularly during critical events like elections, collaboration is key. We see additional opportunities for bipartisan support, as well as private and public sector partnerships:

At Resemble AI, we know firsthand that deepfake detection technology is a powerful tool, capable of providing crucial context and labeling to identify potentially misleading or AI-generated content. By distributing and seamlessly integrating this type of technology into widely-used platforms, as seen in today's phone spam filters, we can protect more Americans regardless of their own resources or knowledge in this area.

We've also seen the effectiveness of red team exercises, that have a sole purpose to test the robustness of AI systems by simulating real-world cyber threats and attacks.The red team can create a deepfake audio, video, or image and distribute it within the organization. The blue team, who is working toward improving an organization's security, is then tasked with detecting the deepfake and responding appropriately. This will help assess the effectiveness of the organization's detection and response systems.

Moreover, we suggest the creation of a national task force composed of experts from the private and public sectors, including representatives from AI companies, government agencies, academic institutions, and civil society organizations. This task force would be responsible for developing best practices and standards for the use of AI in election-related content, as well as coordinating efforts to detect and counter AI-generated misinformation.

To further strengthen security measures, Resemble AI supports the implementation of strict penalties for the malicious use of AI in election-related content. This could include fines, legal action, and the revocation of broadcasting licenses for media outlets that knowingly distribute AI-generated misinformation. By establishing clear consequences for the misuse of AI, we can

deter bad actors and create a stronger incentive for compliance with transparency and labeling requirements.

3. **Integrity Verification**

We believe that AI watermarking technology is a readily available solution to verify the integrity of audio content in this current election year, ensuring that voters can distinguish between voices created with consent and unauthorized deepfakes.

We strongly advocate for the implementation of AI watermarking technology, such as our Neural Speech Watermarker, to verify the integrity of audio content in elections. We propose that all election-related audio content, including political advertisements, campaign messages, and public statements by candidates, be watermarked using this technology. The watermark would serve as a tamper-evident seal, allowing voters and election officials to easily verify the authenticity of the content.

To facilitate the widespread adoption of AI watermarking, we recommend the establishment of a certification program for AI watermarking technologies. This program would set standards for the effectiveness and reliability of watermarking solutions, ensuring that only trusted and vetted technologies are used in the election process. Additionally, we propose the creation of a centralized database where watermarked election-related audio content is registered and can be easily accessed by the public for verification purposes.

Furthermore, we recommend the development of user-friendly tools and platforms that allow voters to easily check the authenticity of election-related audio content. This could include a mobile app or website where users can upload suspicious audio files and receive an instant verification of their authenticity. By empowering voters to take an active role in verifying the integrity of the content they encounter, we can create a more resilient and trustworthy election ecosystem.

To ensure the effectiveness of AI watermarking in preventing deepfakes, we propose regular security audits and stress tests of watermarking technologies. These assessments would be conducted by independent security experts and would simulate real-world scenarios to identify potential vulnerabilities and areas for improvement. By continuously evaluating and strengthening the integrity verification measures in place, we can stay ahead of the evolving tactics of malicious actors and maintain the highest standards of security.

Conclusion

Thank you for the opportunity to provide insight into voice cloning technology and preventative measures that can be taken now to ensure the integrity of this year's election. We are always willing to help facilitate partnership between the private and public sectors to ensure today's innovation is used responsibly.

Chairman Blumenthal, ranking member Hawley, Senator Klobuchar, and committee members: I thank you for your stated concerns on the impact deepfakes have on our elections, on our democracy, and on society, and I thank you for holding this hearing, as well as requesting my presence here. It is an honor to provide any insights that may be of help to both your committee and the American people, and I applaud the committee's great efforts in surfacing the deepfake problem in front of the nation.

Over the last two years, there has been an air of excitement and wonder over the many rapid technological advancements and developments in what we now call generative artificial intelligence. Such advancements are now very much a permanent fixture in society, in conversation, and at work, and are slowly becoming just as prevalent as other technologies that were once novel but are now commonplace in our everyday lives. As a longtime cybersecurity official, myself and my team at Reality Defender foresaw the emergence of such technologies into the mainstream well before this happened, and were demonstrably early in working to thwart them before they were an equally mainstream issue. In fact, we built our company because we've seen the tangible impacts that weaponized content and disinformation had on our loved ones, and we sought to stop what are now the most advanced disinformation-driven threats of our time.

Generative artificial intelligence, or Generative AI, is not a fad, nor is it an entirely negative presence in our world. AI, and to a lesser extent generative AI, has been used for many years, largely in the background without the fervor and publicity it gets today. Thanks to the rapid adoption and iteration over the last several years, led largely by the unmatched innovation in American tech companies, we're now privy to concepts that were once works of science fiction.

Myself, my team at Reality Defender, many in the tech world, and many consumers do not feel that *all* generative AI technologies are bad. Far from it. Many uses of this technology can echo the technology that came before it, augmenting human capabilities and increasing the speed of innovation. A hyper-realistic avatar could, for instance, make video calling possible to parts of the country where high-speed internet access is sorely lacking, reducing the need for bandwidth. A chatbot trained on fifth-grade math can not only help a student with their homework, but help them understand how to solve a problem.

There are a million and one incredible and groundbreaking uses for these technologies that will change innumerable lives.

Yet just as generative AI has its immense benefits, so too does it imperil core societal tenets that we hold dear. Deepfakes, the colloquial term for fully or partially AI-manipulated works that seek to mimic existing individuals or create new ones wholesale, are especially troubling in how they can fabricate a new reality for a person or millions of people.

We've seen this on the large scale in recent weeks, with Russian media falsely depicting Ukrainian forces as the perpetrators of the devastating attack in a Moscow music hall, or with the robocall of a fake President Biden to thousands of New Hampshire constituents. This is a

new and precise weapon of propaganda during elections and during wartime, and we are just beginning to see what it can do.

This is a new problem and dangerous problem that is increasing in size, scope, and damage. Our adversaries are getting smarter about how to apply deepfakes effectively to manipulate elections, spread anti-American sentiment, and sow discord.

This is not just a centralized weapon. We see the harms deepfakes, created by a vengeful individual on a smaller scale practically every week, with non-consensual deepfake pornography — over 99% of which depicts girls and women — ruining countless lives of schoolchildren, of notable names, of wives, and of daughters. Even high school students in Baltimore, MD and Putnam County, N.Y. managed to degrade the reputation of their school principal with a high quality deepfake.

I cannot sit here and list every single malicious and dangerous use of deepfakes that has been unleashed on Americans, nor can I name the many ways in which they *can* negatively impact the world and erode major facets of society and democracy. I am, after all, only allotted five minutes. What I can do is sound the alarm on the impacts deepfakes can have not just on democracy, but America as a whole.

Generative AI models and technologies are available to anyone with basic internet access and either a credit card, or, as is often the case, for no money at all. Anyone can create an AI-generated audio, video, image, or text file that, by and large, is so convincing that even our own experts with PhDs cannot manually tell if they are real or fake. What's just as important is how anyone can *distribute* this media on a massive scale thanks in part to the popularity of social media and content platforms. These platforms have all but eradicated their trust and safety teams who would normally prevent abuse and disinformation from propagating to millions. Though these platforms have, with minimal effort, kindly asked their users to self-label content, there are no penalties nor further action if they don't.

Because of the democratization of these technologies, because everyone has access to them, and everyone has access to all publicly available materials on the internet, anyone can deepfake anyone without their consent and spread these deepfakes like wildfire. By anyone, I mean men and women, children and adults, Democrats and Republicans. This is, by and large, a non-partisan issue, and one that both parties should equally fear.

There have been countless attempts around the world — both successful and otherwise — to subvert elections and sway public opinion. This year we saw our Chinese and Russian adversaries test out early deepfake-fueled election interference campaigns in Taiwan and Slovenia, respectively. While these efforts may not have changed the overall outcome of the election, they are a warning we must heed and learn from to protect our own elections.

As these materials appeared and spread to millions in minutes, the research and responses in pinpointing them as deepfakes took hours, if not days to spread to a disproportionately smaller

group. As Mark Twain once said, “A lie can travel halfway around the world while the truth is putting on its shoes.” By the time a deepfake has spread like wildfire, any reporting that calls it a deepfake is already too late and will reach only a tiny fraction of those who first saw it. Determining truth will always be slower and more difficult than spreading a lie.

Now, Taiwan and Slovenia are not America, but the same thing can and has, in election materials, in robocalls, and in campaign affiliate-disseminated memes, happened here. This is not fear mongering, nor is it AI alarmism, doomerism, or conspiracy-minded hyperbole. It is simply the logical progression of the weaponization of deepfakes.

If legislative measures are not expeditiously taken to enforce the proactive and detection of deepfakes where Americans consume content, deepfakes will have a sizable impact on our election this November and every election going forward, with potential negative implications and effects on and for both sides of the aisle.

I applaud the efforts of ranking member Hawley and Senator Klobuchar on their Protect Elections from Deceptive AI Act. Unlike previous measures that have more or less given the pen to the largest content and social platforms, this law could meet any potential threats that deepfakes pose to our elections head-on and have tangible consequences for those seeking to disrupt our democracy with deepfakes — as well as the platforms used to disseminate such materials. Though states like Minnesota have passed measures against specific uses of deepfakes, there are no federal laws regarding the detection or prevention of dangerous deepfakes on the books, and this law would do a world of good in ensuring some uses of deepfakes do not, in fact, impact the coming elections and all that come after.

That said, generative AI moves faster than any technology that preceded it. Legislation needs to not only move at the same speed, but faster, forecasting and potentially anticipating ill-intended consequences that may stem down the line from the latest generative AI technologies, all built by companies who "move fast and break things."

Unlike the apps and startups who followed such a motto before them, the "things" in this instance are aspects of society everyone in this room holds dear. Democracy. Truth. Trusting what you see and hear. It is not a stretch to say these are at stake when any American — any person — can both create mis- or disinformation to convince any other person they are anybody, saying anything, in real time, for free or cheap, and with little to no technical expertise needed.

I am here to say that we must treat deepfakes with equal or greater importance as we do with the worst kinds of content that existed before it, precisely because it gets at the heart of what makes us human and wholly erodes the trust of our political dialogue. I urge our elected officials to mandate the removal and expulsion of such dangerous deepfakes and AI-generated media with complete and total urgency and not, as it currently exists, have social platforms and businesses relegate it to yet another chore and decision for the average American to determine and make. We must mutually agree, as Democrats and Republicans, men and women, that the

deepfake threat is a threat to everyone, to the things we hold dear, and to America itself. And we must act quickly and at the same speed that AI progresses, lest we be taken by surprise by new attacks from afar, from at home, and, most importantly, on truth.

Written Testimony of Rijul Gupta

Founder and CEO, Deep Media, Inc.

Senate Committee on the Judiciary, Subcommittee on Privacy, Technology, and the Law

Hearing on "Oversight of AI: Election Deepfakes"

April 16, 2024

Introduction

Chairman Blumenthal, Ranking Member Hawley, and distinguished members of the Subcommittee, thank you for the opportunity to testify today on the critical issue of AI-generated Deepfakes and their potential impact on U.S. elections. My name is Rijul Gupta, and I am the Founder and CEO of Deep Media, Inc., a leading Deepfake Detection and AI Security company at the forefront of combating the threats posed by unethical AI and Deepfake misinformation.

As an expert in the field with over 15 years of experience developing cutting-edge AI algorithms, I am **deeply** concerned about the rapid proliferation of Deepfakes and the significant risks they pose to our democracy. Today, I will provide an overview of Deepfake technology, discuss its potential impact on elections, and propose solutions to address this growing threat.

Defining Deepfakes

Deepfakes are images, audio, or videos that have been generated or manipulated by AI in a manner that may harm or mislead the public. It is important to note that Deepfakes do not include AI-generated text, such as the output of language models like ChatGPT. While

AI-generated text is a related technology, it poses distinct challenges and should be addressed separately.

The human brain is uniquely susceptible to manipulation by AI-generated images, audio, and video in ways that it cannot be influenced by text alone. Deepfake images, audio, and video present a distortion of truth, they present themselves as reality, and in doing so **undermine** truth in a powerful and dangerous way. Deepfakes represent a form of "Counterfeit Truth" that can distort our perception of reality and undermine trust in the very information we consume.

Technology Overview

Modern Deepfake generators employ a combination of advanced AI techniques, including utilizing Transformers, Generative Adversarial Networks (GANs), and Diffusion Models. These models are trained on vast datasets containing **millions** images, audios, and videos including millions of unique human identities and can process hundreds of thousands of samples in just a few hours, enabling them to create highly realistic imitations of human motion and appearance.

In addition to these core AI technologies, Deepfakes often incorporate traditional post-production techniques such as Photoshop, After Effects, and sound design to further enhance their realism, making them nearly indistinguishable to the human eye.

While "Shallowfakes" – simple Deepfake manipulations like filters or image generation – may seem less threatening, they pose a unique and significant risk due to their scale and accessibility. Shallowfakes are highly effective at deceiving viewers on low-resolution phone screens or in

areas with poor internet connectivity, and can be created by anyone with minimal to no training. In concert with higher quality Deepfakes, which remain particularly dangerous because of their near-perfect quality, it becomes increasingly difficult to distinguish manipulated from genuine content. Both Shallowfakes and Deepfakes present unique challenges, and the gap between them is rapidly narrowing as the technology advances, making it crucial to address the **full spectrum** of AI-generated media.

Proliferation of Deepfakes on Social Media

The rapid advancement of Deepfake technology has led to an **exponential increase** in the realism, affordability, and prevalence of Deepfakes on social media platforms. Research and development efforts in the AI industry are continually making Deepfakes more realistic and cheaper to produce, with these advancements typically taking 6-12 months to reach end-users.

Given the substantial investments in AI research over the past six months, we anticipate major problems arising from Deepfakes in the next 3-6 months. Election seasons worldwide have already sparked a significant increase in Deepfake creation and interest, particularly from malicious actors. Disturbingly, many pornographic or unsafe Deepfake models are freely available to **anyone** with an internet connection.

Impact of Deepfakes on US and Global Elections

While the United States is not unique in facing the challenges posed by Deepfakes, their potential impact on our elections is particularly concerning. Deepfakes can and are used for voter intimidation and confusion, as exemplified by the Biden robocall incident. They can also be

employed for virtual political assassination, such as the circulation of fake images depicting Donald Trump's arrest. Deepfakes can redirect narratives, as seen in the fabricated Hillary Clinton endorsement of Ron DeSantis, or create the illusion of groundswell support, like the manipulated images of Trump with Black supporters or Eric Adams speaking Spanish.

However, the most alarming aspect of Deepfakes is their ability to provide bad actors with plausible deniability, allowing them to dismiss genuine content as fake. This erosion of public trust strikes at the very core of our social fabric and the foundations of our democracy. The human brain, wired to believe what it sees and hears, is particularly vulnerable to the deception of Deepfakes. As these technologies become increasingly sophisticated, they threaten to undermine the shared sense of reality that underpins our society, creating a climate of uncertainty and skepticism where citizens are left questioning the veracity of every piece of information they encounter. In a world where the **very nature of truth** is called into question, the foundations of our democratic institutions, which rely on an informed and engaged citizenry, are at risk.

Deepfake Solutions

On the technological front, while adding clear metadata and watermarking manipulated images have served as important first steps in combating Deepfakes, they alone are not sufficient. These techniques can be easily circumvented by bad actors, and as Deepfake technology continues to evolve, more robust and comprehensive solutions will be necessary to effectively detect and mitigate the threat posed by these malicious AI-generated media.

The development of advanced Deepfake detection platforms on the other hand have proven critical, and show great promise for detecting even the highest quality Deepfakes. Just as the models used to create Deepfakes continue to evolve, so too must our detection efforts. Investigative journalists, such as Donnie O'Sullivan at CNN, Geoff Fowler and the Washington Post, and Amanda Florian at Forbes, play a vital role in uncovering and exposing Deepfakes.

Government entities, including the military and intelligence community, must also actively engage in Deepfake detection and mitigation. Independent community organizations like Witness, the DARPA AI Force, the Content Authenticity Initiative, and Deep Media-led coalitions are essential in coordinating research, sharing best practices, and advocating for effective policies.

Concluding Remarks

As we navigate the challenges posed by Deepfakes, it is important to recognize that the growth of Generative AI and the development of appropriate policies and regulations **need not be in conflict**. Our Generative AI companies represent an economic boon in the global competition with China and other near-peer nations and play a crucial role in protecting the US from rapidly proliferating external AI-related threats.

However, Deepfakes represent a clear market failure – an abuse of a public good that creates negative externalities and erodes trust in the information era. By internalizing these negative

outcomes through smart regulation and industry collaboration, we can accelerate the growth of the Generative AI market while ensuring its safety and integrity.

I **believe** in the positive potential of AI, and I am optimistic about the future. The fact that we are holding this hearing today is a testament to our collective commitment to addressing the challenges posed by Deepfakes. By working together to implement effective solutions, we can harness the power of AI to enrich our lives, improve our political discourse, and build a brighter future for all.

Thank you for your attention and the opportunity to contribute to this critical discussion.

NEW HAMPSHIRE SECRETARY OF STATE
David M. Scanlan

Testimony of David Scanlan, New Hampshire Secretary of State

Before the U.S. Senate Committee on the Judiciary, Subcommittee on Privacy, Technology, and the Law

April 16, 2024

Chairperson Blumenthal, Ranking Member Hawley, and Members of the Committee: Thank you for inviting me to testify at today's important hearing.

These are challenging times for secretaries of state, election administrators and voters. Over the past decade, there have been a series of rapid-fire events that have had a significant impact on the conduct of elections.

Prior to and including the 2016 election cycle, misinformation and disinformation were on the rise along with an increasing decline in faith and confidence in the election outcomes. Social media was a growing source of election-related information.

In 2018, cybersecurity was the major concern. Intrusion into electronic voting systems and databases by foreign actors and malicious individuals was the great fear. The federal government designated these state-run systems as critical infrastructure and appropriated funding to the states through the Help America Vote Act (HAVA) to secure and harden the states' election systems.

The pandemic arrived before the 2020 elections and forced the states to be innovative in finding ways to conduct an election in the middle of a public health crisis.

A dramatic increase in public scrutiny of the election process along with renewed concerns over foreign actor intrusion and misinformation and disinformation were the primary issues in 2022.

Now in 2024, we are faced with artificial intelligence (AI) that can create convincing election related deepfakes, and New Hampshire may be the first state to have been challenged by this technology during its January 23rd Presidential Primary.

The New Hampshire Robocall Using an AI Deepfake

On January 21, 2024, just two days before the New Hampshire Presidential Primary, the NH Attorney General and Secretary of State began getting complaints from voters about

107 North Main St., Concord, NH 03301
(603) 271-3242 | elections@sos.nh.gov

a telephone robocall they received with President Biden's voice asking the voter not to participate in the upcoming Primary Election and to save their vote for the General Election in November when it would be more important. The message was clearly meant to deceive and was a form of voter suppression.

Interestingly, the robocall also "spoofed" a prominent NH democrat and former state party chair by linking a phone number associated with her on the caller ID. Voters began calling her asking for clarification, and she forwarded the complaints to the Attorney General, who, in turn, opened a criminal investigation.

Given the proximity of this event to the Presidential Primary, the state and national media were on top of the story. The Secretary of State, Attorney General and the Governor were able to promptly deliver a unified message alerting voters of the AI-generated robocall, informing them that it was an illegal attempt to confuse voters about their participation in the Presidential Primary.

It is believed that somewhere between 5,000 and 25,000 democratic voters received the robocall. While it is hard to know if any voters declined to vote because of the robocall, New Hampshire did experience a record voter turnout in a Democratic Presidential Primary when an incumbent president was running for a second term.

After the election, the Attorney General quickly tracked down the source of the robocall, and the company utilized to generate the fake voice of the President using AI. The criminal investigation is still pending.

While the crime was an attempt at voter suppression, AI was only a tool to accomplish the goal. A live voice impersonator could have accomplished the same thing. What is concerning was the ease with which the suspect was able to use AI to generate a deceptive message. Imagine adding a video component to the robocall to create an image of the target person making a statement or engaging in a compromising act that, in fact, never happened. The question then becomes at what point does the AI-generated presentation cross the line from being parody or satire protected under free speech to something more malicious and illegal?

NH Legislation Addressing AI Election Deepfakes

House Bill 1596 is currently making its way through the New Hampshire Legislature. The bill would statutorily define terms such as artificial intelligence, deepfake, generative AI and synthetic media. Deepfakes targeting candidates for office would be prohibited within 90 days of an election unless a prominent disclosure appears with the deepfake communication. While not currently a provision in this bill, election officials should be similarly protected to help maintain their credibility as individuals who maintain free and fair elections.

The Broader Picture

Secretaries of State and elections administrators have been challenged with addressing misinformation and disinformation related to the conduct of elections. There are many potential sources of misinformation including foreign actors, opposing candidates attacking each other, or special interest groups manufacturing a crisis or issue to raise money and generate support. The National Association of Secretaries of State organized a campaign that has been in place now for several election cycles promoting election officials as the "Trusted Source" of election-related information. In New Hampshire, we encourage voters to contact their local or state election officials for accurate information on election processes. We accomplish this with a heavy social media presence and with the generous help of news media outlets.

After becoming New Hampshire Secretary of State in January of 2022, I created the Special Committee on Voter Confidence made up of a diverse group of politically accomplished New Hampshire citizens from the two major political parties. They travelled around the state hearing presentations from party leaders, academics, pollsters, and poll workers. Most importantly, each meeting was open to input from the public. The public testimony received was civil and reflected the full range of political viewpoints. There was a general appreciation from the voters that they had an opportunity to speak to the government through a sounding board. The exercise was productive, and it was clear that despite our polarized electoral politics, there is common ground in areas like training poll workers and post-election audits of ballot counting devices.

My basic take-away from the Special Committee on Voter Confidence is that we need to make the election process as transparent as possible, and we need to do a much better job of educating the voting population on how our elections are run, informing them of the many checks and balances at work in every polling place.

Final Thoughts

Attempts to impact the outcome of elections through misinformation and disinformation are nothing new. We experience this all the time when hotly contested political races resort to negative campaigns. However, over time, the tactics change, and rapidly changing technology that is easily accessible is our new reality. AI deepfakes are just the latest significant challenge facing election administrators.

Malicious and illegal attempts to suppress the vote or manipulate the outcome of an election need to be quickly recognized, stopped, and prosecuted.

Any messaging targeting voters that contain AI-generated content should include a disclosure, so voters know exactly what they're looking at. The ability to respond quickly to false messaging is necessary to temper the impact on voters and election outcomes.

Finally, we must focus resources on voter education to help voters recognize when they may be receiving inaccurate election information and let them know where they can obtain trusted and accurate information related to elections.

Senate Judiciary Subcommittee on
Privacy, Technology, and the Law
Hearing on
"Oversight of AI: Election Deepfakes"
April 16, 2024
Questions for the Record
Senator Amy Klobuchar

For Zohaib Ahmed, CEO and Co-Founder, ResembleAI
Some companies are taking a step in the right direction with voluntary commitments to regulate AI in political ads, and some states have passed laws in this area too, but more must be done to ensure our laws can keep up with this changing technology.

- Do you agree that voters deserve full transparency, and that we cannot rely on the voluntary commitments of companies or a patchwork of state laws to keep up with this threat?

Ensuring transparency for voters is essential in combating misinformation. We firmly believe that voter education initiatives play a pivotal role in fostering transparency.

Resemble AI supports legislation that requires clear labeling on AI-generated content in the election process. Similarly to how disclaimers appear at the end of political ads, consumers should be made aware that they are interacting with an AI model or AI-generated content.

In fact, Resemble AI makes our free, real-time Deepfake Detector tool (detect.resemble.ai) available for anyone to combat fraud and promote the responsible use of generative voice technologies. With this tool, anyone can quickly verify the authenticity of widely circulated audio content, making it a valuable asset for journalists, content creators, and the general public who are often on the frontline of combating misinformation.

We also propose the creation of a public database where all AI-generated election content is registered, allowing voters to easily access information about the origin and nature of the content they encounter, which should lead to more transparency.

To take it one step further, we recommend the development of public awareness campaigns that inform voters about the existence and potential impact of AI-generated content in elections. These campaigns should provide clear examples of how AI can be used in election-related communications, both positively and negatively, and equip voters with the tools to critically evaluate the information they receive. Language translation is an example of how voice cloning technology can be used responsibly in an election year.

Senate Judiciary Subcommittee on
Privacy, Technology, and the Law
Hearing on
"Oversight of AI: Election Deepfakes"
April 16, 2024
Questions for the Record
Senator Amy
Klobuchar
Rijul Gupta, CEO, Deep Media

Can you expand on your testimony about the risks that AI-generated Deepfakes pose to our elections, including how watermarking alone is not sufficient?

Deep Media Response:

Thank you for the opportunity to expand upon Rijul's testimony regarding the risks that AI-generated Deepfakes pose to our elections and the importance of a multi-layered defense strategy.

Watermarking is a valuable and powerful tool, but it can be circumvented in several ways. Deepfake algorithms can be trained to identify and remove watermarks, rendering them ineffective. Additionally, if a watermark is not embedded robustly enough, it may be lost during compression or other transformations that the media undergoes when shared online. Furthermore, watermarks are only effective if they are consistently applied to authentic media, which requires widespread adoption and standardization. For these reasons, Deep Media believes only a robust Defense in Depth approach can overcome the shortcomings of individual mitigation solutions.

Another powerful tool for detecting AI-generated Deepfakes is metadata analysis. While useful, it is not foolproof because metadata can be easily manipulated or stripped from an image or video file. Deepfake creators can intentionally alter or remove metadata to hide their tracks, making it appear as though the manipulated content is authentic. Moreover, the absence of metadata does not necessarily indicate a Deepfake, as some genuine media may lack metadata due to how it was captured or processed.

Lastly, cryptographic hashing provides another piece of the puzzle by creating unique fingerprints for media verification, but it is not a one-size-fits-all solution. While hashing can

detect changes to a file, it cannot determine the nature or intent of those changes. A manipulated Deepfake and an authentic video that has been slightly compressed or resized may produce different hashes, making it difficult to distinguish between the two. Moreover, hashing is only effective if there is a trusted database of hashes for authentic media to compare against, which requires significant coordination and maintenance.

Recent discoveries confirming China and Iran's Deepfake influence targeting the US 2020 election highlight the importance of proactive measures required for our democratic integrity. From misleading campaign videos, fake endorsements, and disinformation campaigns to fraudulent speeches, manipulated debate footage, and viral fake news, Deepfakes can manipulate public opinion, damage candidate reputations, and sow confusion among voters.

Additionally, Deepfakes can sabotage voter trust, cast doubt on electoral integrity, and even lead to identity theft or targeted manipulation of swing voters. As these scenarios illustrate, the impact of Deepfakes on elections cannot be underestimated, necessitating ongoing vigilance and the implementation of effective measures to detect and combat their influence. Identifying key characteristics of Deepfakes is critical to developing and strengthening our defensive posture.

This is why we believe that a comprehensive, Defense in Depth strategy for Deepfake Detection is necessary. By combining these various techniques in concert with each other, we believe we can create a more resilient defense that addresses the limitations of each individual technique.

By employing these strategies in a cohesive manner, we can create a robust, adaptable defense against the ever-evolving threat of Deepfakes to our electoral process. It is only through this multi-layered approach that we can effectively safeguard the integrity of our elections from the corrosive influence of AI-generated disinformation.

LEAGUE OF WOMEN VOTERS® OF THE UNITED STATES

Statement for the League of Women Voters of the United States
US Senate Subcommittee on Privacy, Technology, and the Law
April 16, 2024

On behalf of our 500,000 members and supporters, the League of Women Voters of the United States (LWVUS) urges the Committee to take action to curb the negative impacts of Artificial Intelligence (AI) in our elections, including by combatting deepfakes. Deliberately false AI-generated political content often attempts to garner support for or opposition to particular candidates or cause confusion around a person's voting time, place, or manner. The League is becoming increasingly concerned that AI-generated disinformation will undermine the role of voters and corrupt the election process.

LWVUS was founded in 1920, working on the front lines of voter education to assist newly enfranchised women in casting their ballots following the ratification of the 19th Amendment. For over a century, LWVUS has remained committed to our mission to empower voters and defend democracy. The League focuses on advocacy, education, litigation, and organizing with our grassroots network of more than a half-million members and supporters across over 750 Leagues in all fifty states and the District of Columbia. The League is nonpartisan — neither supporting nor opposing candidates or political parties at any level of government — and is committed to protecting the freedom to vote.

The League derives our policy positions based on grassroots member support and consensus. As stated in our position on a citizen's right to know and citizen participation:

> *The League of Women Voters of the United States believes that democratic government depends upon informed and active participation at all levels of government. The League further believes that governmental bodies must protect the citizen's right to know by giving adequate notice of proposed actions, holding open meetings, and making public records accessible.*

Additionally, as stated in our position on campaign finance:

> *The League of Women Voters of the United States believes that the methods of financing political campaigns should provide voters sufficient information about candidates and campaign issues to make informed choices; [and] ensure transparency and the public's right to know who is using money to influence elections.*

These positions are applicable to the issue of deceptive AI campaign communications and deepfakes. Voters deserve access to true, genuine, and complete information about elections and the

candidates seeking their votes. The distribution of disinformation, especially online, has been used in recent elections to sow polarization and distrust in election results in our country. It is crucial that we address the avenues of mis- and disinformation that circulate around an election.

As an organization dedicated to empowering voters, we work to simplify the voting process and make voting accessible, breaking down barriers to participation. Our Democracy Truth Project aims to strengthen democracy and restore trust in the electoral process by combating mis- and disinformation. Through our Democracy Truth Project, Leagues have been trained to engage and report unusual social media and AI-supported election-based misinformation through our partnership with the Algorithmic Transparency Institute (ATI) to expand the Civic Listening Corps (CLC).

The CLC enables the League to contribute to a national mis- and dis-information data set, providing trends across organizations. It allows for state and local Leagues to report to our national office to ensure we are looking at and identifying trends of mis- and disinformation and how best to rebut that information.

Our partnership with ATI, part of the National Conference on Citizenship, combines their technical expertise with the League's 'people power' to identify, analyze, report back, and take action on the latest disinformation trends.

In its first year (2023), the Democracy Truth Project cohort hosted over three hundred meetings with election officials, forty-seven training courses on mis- and disinformation, and 265 events focused on the election process.

By collaborating with stakeholders across sectors and harnessing the power of litigation, communications, and grassroots organizing, we aim to fortify our democratic institutions against the emerging threats posed by AI-driven disinformation campaigns. However, it should not fall solely to organizations such as the League to provide verified, reputable information to voters and ensure transparency in our election process. Congress should similarly break down barriers to participation by reducing the influx of election mis- and disinformation.

The 2024 election is just months away, and primary season revealed some of the deepfake tactics being used to manipulate voters. Two days before the New Hampshire presidential primary, robocalls were sent to New Hampshire voters with a deepfake, simulated voice of President Joe Biden discouraging them from participating in the primary. The New Hampshire robocalls urged recipients not to vote in the primary and to "save" their vote for the November 2024 US Presidential Election.

In response, the League of Women Voters of New Hampshire, the League of Women Voters of the United States, and individual New Hampshire voters filed a federal lawsuit against Steve Kramer,

Lingo Telecom, LLC, and Life Corporation for voter intimidation, coercion, and deception ahead of the presidential primary. This litigation underscores the critical need for proactive measures against malicious actors seeking to undermine the integrity of our democratic processes. Bad actors' utilization of AI-generated content impersonating political figures constitutes a grave threat to voter confidence and participation.

The example in New Hampshire is one of many instances of AI-generated disinformation this year. Fake and deceptive content has been used to target voters of all political party affiliations and has been condemned as unfair and undemocratic by elected leaders across the political spectrum. AI companies have said their tools should not be used in political campaigns, but enforcement has been spotty. The incomplete framework of regulation on AI leaves us vulnerable to efforts to undermine our elections — both from domestic and foreign entities. It is imperative that Congress curb bad actors' ability to use AI in our elections.

Voters deserve to know that the political advertisements they see and hear are free of misleading information or fraudulent misrepresentation. The League hopes to be a resource and partner in this endeavor to ensure that our elections are free and fair and that voters can make informed voting decisions. If you have questions or would like to discuss this further, please contact Jessica Jones Capparell, Director of Government Affairs at JJones@lwv.org.

215 Pennsylvania Avenue, SE • Washington, D.C. 20003 • 202/546-4996 • www.citizen.org

April 16, 2024

United States Senate
Senate Judiciary Committee
Subcommittee on Privacy, Technology and the Law
Dirksen Senate Office Building, Room 226
Washington, D.C. 20510

Submitted electronically

RE: Public Citizen Applauds the Senate Judiciary Committee for Proceeding with Hearings on Deepfakes in Election Communications

Public Citizen submits the following comments in support of Senate legislation to regulate deepfakes in election communications:

So, you think large sectors of the American electorate have been fooled by misinformation and deceptive campaign ads in the past? Just wait until 2024.

This time around, many voters won't know what is true and what is false in the television, radio and social media campaign ads forced upon us by campaigns.

Due to rapid advances in artificial intelligence (A.I.), campaigns are already running ads that look and sound like actual candidates and events, but in fact are entirely fabricated by computer technology. These A.I.-generated ads look and sound so real that it is becoming exceedingly difficult to discern fact from fiction.

When A.I.-generated content depicts a candidate saying or doing things they never did – deliberately designed to damage that candidate's reputation or to deceive voters – these ads are known as "deepfakes." The practice is currently legal in federal elections and all but a dozen states. These ads are not even subject to a disclaimer noting that the content never happened in real life.

Immediately following President Joe Biden's announcement that he is running for reelection in 2024, the RNC produced its first entirely fabricated A.I. campaign ad. It pictured Biden and Vice President Kamala Harris laughing at their reelection party, then spanned into images of China bombing Taiwan, then pictured a collapsing Wall Street financial market, switched to films of 80,000 illegal immigrants flooding across the border, and finally showed a police occupation of San Francisco due to an out-of-control fentanyl epidemic.

All of it was fabricated, but many viewers thought some of it was real, even though the ad included a disclaimer (to the RNC's credit).

Republican presidential candidate Ron DeSantis posted A.I.-generated images of former President Donald Trump having a friendly visit and hugging controversial health official Dr. Anthony Fauci. It never happened. Trump responded with his own A.I.-generated ad of DeSantis enjoying the company of Elon Musk and George Soros. Obviously, this also never happened.

On the eve of Chicago's most recent city election, mayoral candidate Paul Vallas was depicted in an A.I.-generated social media ad, in a voice identical to his own, condoning police brutality.

In an unregulated political environment, expect more deepfakes – many more. In the waning days before the election, it is reasonable to assume candidates, political parties, and especially outside groups will feel free to air deliberately deceptive deepfakes that depict opponents partying in an orgy, praising hostile foreign nations, and committing an assortment of felonies – all looking and sounding real.

Public Citizen filed an amended petition for rulemaking with the Federal Election Commission (FEC) to regulate deepfakes under the "fraudulent misrepresentation" law, which prohibits candidates from fraudulently claiming that opponents said or did something in a way that would damage their reputations when they in fact did no such thing. The agency received 2,400 public comments, most comments encouraged the FEC to proceed with the rulemaking proposal. The Commission will decide whether to formalize regulations of A.I.-content in campaign ads sometime this summer, which will likely be too late to affect the deepfakes in the 2024 election.

In the meantime, legislative action is also urgently needed.

The "fraudulent misrepresentation" law does not apply to outside groups, which have far greater incentives than candidates to abuse A.I. technology. Congress is finally picking up the slack. Rep. Yvette Clarke (D-NY) introduced legislation to require "clear and conspicuous" disclosure of A.I.-content in campaign ads. Sen. Amy Klobuchar (D-MN) has gone a step further and introduced several bills, one of which would ban deepfakes in campaign ads immediately before an election. Another would require disclosure that the A.I.-content is not real. It is getting late but Congress still has an opportunity to pick up these bills in time for the 2024 elections.

At the same time the states are leading the way. Public Citizen is promoting and tracking state action on this issue. We have drafted a model state law, and update a state tracker of legislation (https://www.citizen.org/article/tracker-legislation-on-deepfakes-in-elections/) So far California, Idaho, Minnesota, Michigan, New Mexico, Utah, Washington and Wisconsin have laws on the books. Most of these laws require disclosure of deepfakes or A.I.-content in ads; others outright ban deepfakes near an election. Often these restrictions apply to candidates for any elective office on their state ballots, including federal candidates. If more key states, such as New York, would follow suit that would put a big crimp in state and national deepfake campaign ads.

The price to democracy may be dear. Many voters are already cynical of elections. If voters face a tsunami of real-looking news-like ads that feed entirely false stories about candidates, the public's confidence in the integrity of elections will further be in peril.

Public Citizen strongly encourages swift congressional action to regulate A.I.-content in political communications, before the likely deluge of deepfakes coming in the 2024 elections.

Sincerely,

Craig Holman, Ph.D.
Public Citizen
215 Pennsylvania Avenue S.E.
Washington, D.C. 20003
(202) 454-5182
cholman@citizen.org

○