S. HRG. 118–40

ARTIFICIAL INTELLIGENCE AND HUMAN RIGHTS

HEARING

BEFORE THE

SUBCOMMITTEE ON HUMAN RIGHTS AND THE LAW

OF THE

COMMITTEE ON THE JUDICIARY
UNITED STATES SENATE

ONE HUNDRED EIGHTEENTH CONGRESS

FIRST SESSION

JUNE 13, 2023

Serial No. J–118–21

Printed for the use of the Committee on the Judiciary

U.S. GOVERNMENT PUBLISHING OFFICE
52–709 PDF WASHINGTON : 2024

COMMITTEE ON THE JUDICIARY

RICHARD J. DURBIN, Illinois, *Chair*

DIANNE FEINSTEIN, California
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

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, *Chief Counsel and Staff Director*
KATHERINE NIKAS, *Republican Chief Counsel and Staff Director*

SUBCOMMITTEE ON HUMAN RIGHTS AND THE LAW

JON OSSOFF, Georgia, *Chair*

DIANNE FEINSTEIN, California
RICHARD BLUMENTHAL, Connecticut
PETER WELCH, Vermont

MARSHA BLACKBURN, Tennessee, *Ranking Member*
JOHN KENNEDY, Louisiana
JOSH HAWLEY, Missouri

SARA SCHAUMBURG, *Democratic Chief Counsel*
KAITLYN LANE, *Republican Chief Counsel*

CONTENTS

JUNE 13, 2023, 2:36 P.M.

STATEMENTS OF COMMITTEE MEMBERS

WITNESSES

ARTIFICIAL INTELLIGENCE AND HUMAN RIGHTS

TUESDAY, JUNE 13, 2023

UNITED STATES SENATE,
SUBCOMMITTEE ON HUMAN RIGHTS AND THE LAW,
COMMITTEE ON THE JUDICIARY,
Washington, DC.

The Subcommittee met, pursuant to notice at 2:36 p.m., in Room 226, Dirksen Senate Office Building, Hon. Jon Ossoff, Chair of the Subcommittee, presiding.

Present: Senators Ossoff [presiding], Blumenthal, Welch, Blackburn, Kennedy, and Hawley.

Also present: Chair Durbin and Senator Padilla.

OPENING STATEMENT OF HON. JON OSSOFF, A U.S. SENATOR FROM THE STATE OF GEORGIA

Chair OSSOFF. The Subcommittee on Human Rights and the Law will come to order. Welcome all to today's hearing. It is great to see a packed house. It demonstrates the intensity of interest in this subject. I want to thank you, Ranking Member Blackburn, for working so hard and so closely with me to develop this important bipartisan hearing. And I want to thank each of our witnesses for your participation today.

Throughout history, transformative technologies have emerged with the potential to disrupt societies, economies, and politics profoundly and sometimes very quickly. Machine learning and artificial intelligence may be such a technology. AI capabilities are growing rapidly and in ways even its creators cannot predict. And already it's changing our lives. American families are now threatened by AI-enabled scams made far more sophisticated through this technology than traditional spam email or sham telemarketing calls.

Today we will hear from Jennifer DeStefano, who was targeted by a scam using a deepfake of her 15-year-old daughter's voice to fake her kidnapping and extort a ransom payment.

AI also has profound implications for civil rights, for the criminal justice system, for our democratic and constitutional processes, and for our privacy. Its potential impact on the future of work could include fundamental shifts in education, in recruitment, candidate screening and hiring, and perhaps even more significantly, rapid disruption of labor markets as certain professions are automated.

This technology has profound implications for the future of warfare, as kill chains are automated and predictive technology influences and mediates competition between nation states. As AI tech-

nology develops, great powers competing in an AI arms race engaged in strategic competition, where AI is influencing the decisions made by leaders and militaries, face a different and new risk of escalation and miscalculation.

And some influential technologists and engineers, including prominent figures and prominent leaders of the industry, warn of existential risks ranging from catastrophic political destabilization to the development and deployment of weapons of mass destruction, to catastrophic cybersecurity threats, and to unforeseeable and unknown forms of risk that may emerge alongside more and more powerful forms of artificial intelligence.

Our study of these technologies and associated risks should not blind us, of course, to this technology's extraordinary potential. For example, cancer diagnoses, the development of new life-saving drugs and therapies, productivity growth, and the new forms of technological innovation that AI itself could help us to unlock.

But at a moment like this, it is imperative that Congress understand the full range of risks and potentials to ensure this technology can be developed, deployed, used, and regulated consistent with our core values, consistent with our national interest, consistent with civil and human rights. So I look forward to a productive conversation with this talented and extraordinary panel this afternoon. And with that, I turn to the Ranking Member of the Subcommittee, my colleague from Tennessee, Senator Blackburn.

OPENING STATEMENT OF HON. MARSHA BLACKBURN, A U.S. SENATOR FROM THE STATE OF TENNESSEE

Senator BLACKBURN. Thank you, Mr. Chairman. I am delighted that we're getting the Subcommittee off the starting blocks today. So I thank you and your team for the good work on those efforts and focusing on something where we do share an interest, which is artificial intelligence and technology, and the uses that you say, for good or for bad.

I do want to touch on China, and I'm so pleased that we're looking at this from the human rights angle. I've watched what has happened in China and how they are using AI to grow the surveillance state. And they're very aggressive in this. And we know that they have used it—a good example is the way they have exploited vulnerabilities in Apple's iPhone in the iMessage system to surveil and track the Uyghur Muslims in Xinjiang Province.

And the CCP uses facial recognition as a part of their tracking, and a part of their data, and the logging of information that they do as they're following people. And we want to dive into that a little bit. We know that China is pushing to win the race on AI. They've been very upfront about this, and they are looking to win the race on other technologies.

Quantum computing, 5G, 6G, anything that they see as groundbreaking that helps them to control environments, situations, and people. I think the data from McKinsey & Company should be something that we all look at and take to heart. They predict that by 2030, China's growth in AI could account for up to $600 billion in economic value. And this is exactly what they want.

And in 2017, the National AI Development Plan that they brought forward, China declared its goal of becoming the world

leader in AI by 2030. And they're pursuing this. They're the most aggressive filer of patents for AI technologies. They are constantly challenging our innovators through the PTAB process.

So we should be watching their goals. And this should concern each and every one of us who cares about preserving the freedoms and the democratic values that we hold here in America. As we work to deploy AI technologies here, we need to make a conscious effort to consider the potential impact that those technologies could have on human rights and on how we approach issues such as data collection, data retention, surveillance, and, of course, deepfakes.

This is not to say that we should halt AI development in its tracks or look at approaches that would regulate it out of existence. To the contrary, doing that would practically guarantee that China becomes the world's leader in AI, giving it the opening that President Xi wants to impose the CCP's authoritarian values around the world. But we do need to think carefully about how we deploy AI technologies in the absence of a national privacy law, which we still do not have, a Federal online consumer privacy protection.

We also need to be careful about how we identify and how we stop unauthorized utilizations of AI, whether to surveil or to scam unsuspecting people. So to our witnesses, thank you for being with us today. Mr. Chairman, I appreciate the hearing. Look forward to moving to questions.

Chair OSSOFF. Thank you, Senator Blackburn. I will now introduce our witnesses, and thank you, again, for joining us today.

Ms. Jennifer DeStefano, mother from Arizona, was the victim of a horrifying scam using an AI-generated deepfake of her daughter's voice to fake her kidnapping and demand a ransom. Ms. DeStefano, I think every parent in America who read your story was chilled to the bone by what you experienced. We'll hear from you, Ms. DeStefano, about your experience to help shed light on how AI is being used to supercharge extortion-based scams and threaten the safety of American families.

Dr. Aleksander Madry is a nationally recognized expert on AI and machine learning whose research focuses on how to ensure AI tools are reliable and well-enough understood to be safely and responsibly deployed in the real world. Thank you, Dr. Madry.

Ms. Alexandra Reeve Givens is the CEO of the Center for Democracy and Technology, which works to ensure emerging technologies protect democratic values and advance human rights.

And Mr. Geoffrey Cain is a Senior Fellow at the Foundation for American Innovation, a technologist and author who studies how repressive governments deploy novel technologies and how democracies can respond and defend human rights.

Thank you all so much for joining. Before your opening statements we will swear in our witnesses. If you would all please rise and raise your right hands?

[Witnesses are sworn in.]

Chair OSSOFF. Let the record reflect the witnesses have responded in the affirmative. You may be seated. And Ms. DeStefano, we'll begin please with your opening statement. You'll see some lights indicating time, but we want to make sure you have time to tell your whole story. So don't worry too much about the clock. We're eager to hear from you. And you may begin.

STATEMENT OF JENNIFER DeSTEFANO, VICTIM OF AI DEEPFAKE KIDNAPPING AND EXTORTION SCAM, SCOTTSDALE, ARIZONA

Ms. DESTEFANO. Thank you so much, Senator. I appreciate that. Good afternoon, Senators. It is my great honor to speak with you today and share my experience on how artificial intelligence is being weaponized to not only invoke fear and terror in the American public but in the global community at large as it capitalizes on and redefines what we have known as familiar.

I would like to take this moment to thank Senator Ossoff for inviting me to be here today, and I'd also like to thank Senator Blackburn for your concern on this ever-evolving topic and community threat. AI is revolutionizing and unraveling the very foundation of our social fabric by creating doubt and fear in what was once never questioned, the sound of a loved one's voice.

What is familiar? How many times have you received a phone call from your child and asked them to verify who is calling? How many times has a loved one reached out to you in despair and you stopped them to validate their identity? Did you hang up on them? Did you require to call them back to make sure you are speaking to the correct person? The answer is, more than likely, never. The sound of a loved one's voice is often never authenticated. It has a unique identity, as unique as a fingerprint. This familiar identity is innate and is designed by God. It is what binds a mother to their child and a newborn infant to their mother.

January 20th was a typical Friday afternoon for our family, kicking off a weekend of races and rehearsals. We often divide our family across the State. It's divide and conquer. My husband was with our older daughter, Brie, training for a ski race, and I was with my younger daughter, Aubrey, picking her up from a rehearsal at dance. Brie had not raced in years and promised me that she would take it easy.

At about 4:53 p.m., I received a call from an unknown number upon exiting my car. At the final ring I chose to answer it, as unknown calls we're very familiar with—can often be a hospital or a doctor. It was Briana sobbing and crying, saying, "Mom?" At first, I thought nothing of it and casually asked her, "What happened?" I had the phone on speaker, walking through the parking lot to meet her sister.

Briana continued with, "Mom, I messed up," crying and sobbing continually. Not thinking twice, I asked her again, "Okay, what happened?" Suddenly, a man's voice barked at her, "Lay down. Put your head back." At that moment, I started to panic. My concern escalated as I demanded to know what was going on. But nothing could have prepared me for her response that she gave me next.

"Mom, these bad men have me. Help me, help me, help me." She begged and pleaded as the phone was taken from her. A threatening and vulgar man took the call over. "Listen here, I have your daughter. You call anybody, you call the police, I'm going to pop her stomach so full of drugs, I'm going to have my way with her, I'm going to drop her in Mexico, and you'll never see your daughter again."

As I had my hand shaking on the door handle of the dance studio, I ran inside and started screaming for help. The next few min-

utes were every parent's worst nightmare. I was fortunate to have a couple of moms there who knew me well, and they instantly went to action.

One mom ran outside and called 911. My younger daughter Aubrey was standing there listening to all the vulgar threats this man was making that he was going to do to her sister. I needed her help and asked her to start calling her dad, call her brothers, call anybody we have to find her sister. She stood there paralyzed in fear.

The second mom ran to Aubrey's aid and started making calls to her dad. The kidnapper demanded a million dollars. That was not possible. So then he decided on $50,000 in cash. That was when the first mom came back in and told me that 911 is very familiar with an AI scam where they can use someone's voice. But I didn't process that. It wasn't just her voice. It was her cries. It was her sobs. It was just not her voice. She said okay and left.

I continued with the negotiations for the ransom. I asked them for wiring instructions, routing numbers, but they refused. Instead, they required me to get in a van with a bag over my head with $50,000 in cash to be transported to my daughter. If I didn't have all the money then we were both going to be dead. I was shocked.

At that point in time the second mom came back to me, and she had located my husband who had found Brie resting safely in bed. She came to me and told me that Briana was safe, but I did not believe her because I had just spoken to my daughter, and I was very sure of her voice, and I was very sure of her cries. So I demanded to talk to my daughter.

Briana got on the phone, and she had no idea what was going on, and she kept reassuring me that she was safe. I asked her so many times, "Are you sure? Are you sure you're safe? Are you sure you're with dad? I spoke to you. How can you be in both places at once?" I asked her over and over again. My mind was whirling.

When I finally had the reassurance I needed, I knew she was safe, and I was furious. I lashed at the men for the horrible attempt to scam and extort money. They continued to threaten to kill Brie. I made a promise that I was going to stop them and they were never going to hurt my daughter nor anybody else again.

At that point, I hung up and collapsed to the floor in tears of relief. I called the police to pursue the matter, and unfortunately, I was met with, "It was a prank call," that it happens often, and that there's nothing that can be done, and that I probably am not in harm's way but it's not a guarantee. They offered to have a police officer contact me, again from an unknown number, as authorities are calling from blocked numbers. But that's all they could offer. That certainly did not make me feel better.

The bottom line was no actual crime had been committed, so no physical kidnapping had taken place and no money had transferred, period. The end. But that wasn't the end. It couldn't be the end. If it was the end then how would this nightmare ever stop? I stayed up all night paralyzed in fear. "Do they know where I am? Do they know where my daughter is? How did they get her voice? How did they get her crying, her sobs that are unique to her?" She is not a very public person. I was wondering, "Are we being cyberstalked? Targeted?"

So many questions that were left unanswered. So I turned to the community, and the responses were overwhelming. Friends and neighbors came out of the woodwork with their stories. Kidnapping, phone calls coming from their children's phones, bags of money being driven halfway to Mexico, even voices of young children nowhere to be found on social media who do not have phones. The stories kept pouring in.

My own mother even received a phone call with my brother's voice, claiming to be in an accident needing money for a hospital bill. The common response that victims received from authorities when reported was that nothing could be done. In fact, one mother I know personally shared with me how she was even mocked by her son's school and a security officer. The caller even used her son's unique nickname to self-identify. Fortunately, he was safe in class, and she was told this happens all the time as her fear was dismissed.

Money scams have been around for thousands of years. This is entirely different. This is terrorizing, lasting trauma. Even months later sharing the story makes me shake to my core. Aubrey was approached by a boy to hang out sometime, and she concluded it was because he wants to kidnap her. That's not a normal 13-year-old thought. It was my daughter's voice. It was her cries. It was her sobs. It was the way she spoke.

I will never be able to shake that voice and the desperate cries for help out of my mind. It's every parent's worst nightmare to hear your child pleading with fear and pain, knowing that they are being harmed and you're helpless. The longer this form of terror remains unpunishable, the farther and more egregious it will become. There is no limit to the depth of evil AI can enable.

The thought crossed my mind before I hung up on the kidnappers to follow through with the physical abduction of me. Was that what it would take to bring this to an end? Was that what it would take in order to have a punishable criminal offense?

As our world moves at a lightning-fast pace, the human element of familiarity that lays foundation to our social fabric of what is known and what is truth is being revolutionized with AI, some for good and some for evil. No longer can we trust, "Seeing is believing," or, "I heard it with my own ears," or even the sound of your own child's voice.

The concept redefines and rewrites what the very meaning of familiarity means. I ask you, when your mother calls are you going to hang up on her and call her back to make sure it's her? When your child calls in need of help will you end the call and say, "I don't believe it's really you"? Is this our new normal? Is this the future we are creating by enabling the abuses of artificial intelligence without consequence and without regulation?

I want to thank you for your time and attention today. Congress has a large and looming task ahead. How do we move forward as a community with this haunting reality that is plaguing us? If left uncontrolled, unregulated, and we are left unprotected without consequence, it will rewrite our understanding and perception of what is and what is not truth. It will erode our sense of familiar as it corrodes our confidence in what is real and what is not. This is a

non-partisan matter, and I've seen the hands reach across the aisle in unified concern.

That gives me great hope. How to contain the ever-evolving artificial intelligence and its unknowns is not an easy task. My sincere thanks and humble appreciation for your time and attention today. I thank all of you, especially Senator Ossoff and the Senate at large, for tirelessly taking action to keep our community and our world safe from the hands of evil. I am one person, one story, but I'm not the only one. And I certainly will not be the last unless action is taken. I wish you Godspeed.

[The prepared statement of Ms. DeStefano appears as a submission for the record.]

Chair OSSOFF. Thank you, Ms. DeStefano, for sharing your powerful and disturbing story. And we will in more detail investigate all of the issues you've raised. I appreciate it. Dr. Madry, it's now your turn for an opening statement. Thank you.

STATEMENT OF ALEKSANDER MADRY, CADENCE DESIGN SYSTEMS, PROFESSOR OF COMPUTING, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE, MASSACHUSETTS

Dr. MADRY. Thank you. Chairman Ossoff, and Ranking Member Blackburn, and Members of the Committee, thank you for inviting me to testify today. I must say it's hard to follow this testimony, in part because honestly it makes way better some of the points I wanted to make, but let me try nonetheless.

So I want to focus my testimony on a single issue that I find particularly salient, time sensitive, and really unsettling: how AI could undermine our whole information ecosystem, and with that erode how our society functions and carries out democratic decision-making.

The newest wave of generative AI is poised to fundamentally transform our collective sense making. And this is due to two reasons. First, AI enables the creation of content that is not only extremely realistic but also persuasive, even though it may be false. Second, with AI, the creation of such content is cheap and broadly accessible, making it frighteningly easy to deploy at scale.

As a result, a whole spectrum of risks is emerging. Firstly, traditional spam, scam, and phishing become even easier to conduct. Also, AI can now convincingly impersonate a human online or over the phone. That was a frightening but very real experience that we just have heard about. So how will our digital platforms cope with swarms of AI-driven bots that can breeze through existing bot detection and moderation algorithms?

And the worst thing is that such boosting of traditional deception is just the beginning, not the end. So AI is now able to create content that is both convincing and personalized. This means that phishing no longer needs to involve generic emails sent out to thousands of recipients. Instead, both the message and the ensuing conversation can be fully automated and customized to you.

AI is also bound to transform how we think about any information campaign, be it ideological, political, or commercial. Such campaigns will no longer need to rely on a promoted message to go viral. Instead such campaigns they can be filtered to generative AI

that reaches an internet audience individually and in a highly personalized manner.

So the hook to get you will not be some post that came across your social media. Rather, it might be a Facebook friend who is actually an AI-driven agent impersonating a human, a friend that only subtly mixes in political commentary or product endorsement into your engaging conversations.

Similarly, campaigning for a cause might no longer require corralling a critical mass of people to do the outreach, be it via direct calling or letter writing. Instead, a single actor could fill such a campaign by themselves, using generative AI-driven bots in place of people. Such a campaign would be equally effective, but it would need neither the buy-in from the broader population nor comparable resources. And as far as I know, this would all probably be legal, too.

Also, AI doesn't just produce content that is personalized. It can also make this content be personable. This could be used to make interacting with AI not only persuasive but also alluring to the point of being addictive. What if these capabilities supercharge the attention economy, or rather distraction economy, that we are having right now? How will we feel about having our children be exposed to that?

Finally, AI is ushering us into the era where any record—a contract, a deposition, a video—could be plausibly faked. How does this affect our collective discourse as well as the legal and governance system? All of these concerns may paint a rather bleak picture, but there's actually much that can be done.

On the technical front, we need tools that help humans judge the extent to which a given content was generated by a human. However, these tools are still developing, and they will not be a panacea. Rather, they can provide the necessary friction that makes it harder to abuse AI capabilities. They will also not work to the full extent, and in some cases at all, without complementary policy developments.

In particular, the efficacy of these technical approaches will hinge on how broadly adopted they are. Policy can accelerate this process. Policies could also require any consumer-facing, AI-generating content to be labeled as such. And policy could also mandate providers of AI services to implement adequate identification and reporting mechanisms.

Finally, we do need to work on AI literacy. I think that no matter what happens in the end, the public needs to understand how to judiciously interact with AI systems and to be on the lookout for when they are actually interacting with AI in the first place. We really do not want to learn this the hard way, the way we have seen it over here.

So to conclude, I am really excited about the positive impacts that AI can have, but we need to be mindful of the very real risks, and we need to get started now. Thank you, and I'm looking forward to the questions.

[The prepared statement of Dr. Madry appears as a submission for the record.]

Chair OSSOFF. Thank you, Dr. Madry. And I would note that the Chair of the Judiciary Committee, Senator Durbin, has arrived. Mr. Chairman, any remarks you'd like to make?

Ms. Givens, your opening statement, please.

STATEMENT OF ALEXANDRA REEVE GIVENS, PRESIDENT AND CHIEF EXECUTIVE OFFICER, CENTER FOR DEMOCRACY AND TECHNOLOGY, WASHINGTON, DC

Ms. GIVENS. Senators, thank you so much for inviting me to testify. I'll say that I spent 5 years of my career sitting on the benches right behind you, so it's a particular honor to be in front of the Judiciary Committee today.

The world's attention is rightly focused on the possibilities and risks of AI systems. As policymakers look to address potential harms and promote responsible innovation, it's essential that they do so with a focus on human rights, including the rights to liberty, privacy, freedom of expression, and equal treatment before the law. My testimony is going to draw us a little broader to focus on two areas where AI systems are already impacting these rights today: the use of face recognition by law enforcement and the impact of generative AI on elections.

In previous testimony I've described how AI systems are also harming people's civil rights and economic mobility, for example, when people are denied a job or housing based on inferences made about them by an AI system or are wrongly accused of fraud because a government agency uses a flawed AI tool. These real-world harms are happening today.

So I hope the key takeaway of today's hearing is this: That at a time when many are discussing the existential risks of AI, there are concrete issues on which Congress and the executive branch can act right now, and in doing so, demonstrate how AI can be governed in a way that centers human rights.

Today, my organization partnered with the Leadership Conference on Civil and Human Rights in over 60 civil society groups, urging the Biden administration to expedite its good work on these issues. These questions impact all sectors of society, and there's much that both Congress and Federal agencies can do.

A few words on AI and government surveillance. Last fall, many of us were inspired by images of the brave protests happening in Iran. But we weren't the only ones watching those protests. In Iran today, face recognition allows the government to identify protesters and take action against them. Face recognition has also been invoked to police women not correctly wearing the hijab, with one official threatening that violators would face immediate penalties, such as their bank accounts being blocked.

In this context, AI systems are enabling a repressive regime to identify dissenters, surveil them, and automate their punishment. AI is used in similar ways in China, as we'll hear, and to intimidate peaceful protesters in Uganda, Hong Kong, and more. Such examples feel far from the U.S., but there have already been abuses here as well. Police in Florida and Maryland have used face recognition to identify and harass peaceful protesters, chilling Americans' free speech and right to peacefully assemble.

Recently a Georgia resident, Randall Reid, was held in jail for 6 days because a face recognition system misidentified him. There are other accounts of wrongful arrests, and these are likely just the tip of the iceberg. My testimony shows recommendations for how Congress could regulate face recognition. But importantly, this is just one area where Congress could draw a clear contrast to autocratic regimes and lead on AI right now.

Turning to my second example, advances in generative AI are spurring creativity and innovation across the country and around the world. But they also raise threats for human rights, including in the context of elections. In past elections, operatives used robocalls and texts to spread deceptive information. But now bad actors could easily use AI to exponentially grow and personalize voter suppression or other targeting.

Generated images can also twist public understanding of political figures and events. Videos and images have already been digitally altered to compromise public officials. Fake content is now cheaper, easier, and more convincing because of the growth of AI tools.

Now, regulating in this space must be approached with care because it involves expressive conduct. There are many legitimate reasons why people use software to generate and alter content, from artists making new works, to parody, to researchers altering celebrity photos to show the hypothetical impact of skin cancer. Barring or heavily restricting such activities would harm free expression and innovation and quickly run afoul of the First Amendment. But this doesn't mean that leaders must sit idle. I'll briefly list four ways in which Congress could act.

First, Congress could require the developers of AI systems that can be used in high-risk settings to disclose how their tools are developed and designed, and require testing for elements such as safety, validity, explainability, non-discrimination, and privacy.

Second, in some instances the appropriate framework to address AI harms will be litigation under existing laws. For example, fraud and extortion, harassment, civil rights, intellectual property, and product liability. Courts are going to have to tackle how these laws apply to new fact patterns, and whether and when AI companies bear liability for the content their tools produce versus downstream users. But Congress can shine a light on these complex issues and act as appropriate to fill in gaps through hearings and reports or an expert commission.

Third, there's an urgent need for AI companies to develop robust safety standards, as CEOs have said themselves in this very building. Governments are pressing companies for near-term voluntary agreements. Congress can help ensure that such agreements are developed with public visibility and engagement from civil society and independent experts.

Fourth and finally, on deepfake specifically, Congress can use its funding and oversight to scale our Nation's capacity at this critical time. This should include supporting the development of detection technologies and ensuring key institutions like law enforcement agencies are equipped to quickly debunk manipulated content. My written testimony shares more on each of these topics. Thank you, and I look forward to your questions.

[The prepared statement of Ms. Givens appears as a submission for the record.]

Chair OSSOFF. Thank you, Ms. Givens. Mr. Cain.

STATEMENT OF GEOFFREY CAIN, SENIOR FELLOW, FOUNDATION FOR AMERICAN INNOVATION, CHICAGO, ILLINOIS

Mr. CAIN. Chairman Ossoff, Ranking Member Blackburn, and Members of the Subcommittee, thank you for the opportunity to testify here today. The Chinese Communist Party, or the CCP, has seized on artificial intelligence to emerge as the greatest threat to democracy and human dignity in the world today.

As an investigative journalist formerly in China, I was among the first people to document and expose the horrific surveillance state that oppressed the Uyghur population in the far western region of Xinjiang. China used its vast AI-powered surveillance system, literally called SkyNet. Since 2017, the atrocity has morphed into the largest internment of ethnic minorities since the Holocaust. The U.S. State Department calls this a genocide.

In December 2017, I was kicked out of China while researching my book, "The Perfect Police State," which is a book about the surveillance dystopia that has been built there. Ever since then, the AI police state has expanded to alarming levels. In 2018, I moved to Turkey and for 3 years tracked down defected former intelligence officers from the Ministry of State Security, the powerful intelligence body in China with a global reach.

These spies from the Ministry told me that the Uyghur genocide was the beginning of an experiment in total AI surveillance. The CCP planned to enlist companies and then expand the experiment nationwide in China and globally wherever possible. In July 2017, China unveiled its National AI Development Plan. It called AI a historic opportunity and pledged to align development with the government's authoritarian values. China has declared its goal as becoming the world leader in AI by 2030.

Recently, the CCP unveiled AI-powered alarms that notify the police when someone unfurls a banner, when a foreign journalist is traveling to certain parts of the country, and when someone from an ethnic minority is present. The companies that helped build China's surveillance apparatus operate here in America.

ByteDance, the megafirm that owns TikTok, the popular social media app, stands accused by a whistleblower of running an in-house CCP committee that had access to all the app's data, including data stored in the U.S., contradicting the company's past testimony at other Committees. This was all according to a court filing.

Other sanctioned AI firms, such as iFlytek, SenseTime, and Megvii have emerged as billion-dollar unicorns with the backing of the Chinese state and the involvement of American venture capital firms. In April, the Cyberspace Administration of China, a very powerful body in the country, announced draft regulations for generative AI as well. These draft rules would require content produced by chatbots to follow, quote, "socialist core values" and avoid information that undermines, quote, "state unity."

Given the CCP's enormous success at censorship so far, I believe that it will once again succeed at coercing and coopting Chinese and American technology firms. It will transform generative AI

into a tool of state oppression. We must abandon the misguided idealism of working with AI companies and government institutes in China.

As long as the CCP has any control over these technologies, AI will not open the democratic discourse, and it will not contribute to the betterment of humanity. China cannot be trusted to help build the guardrails for AI, which is something that Sam Altman, the CEO of OpenAI, recently proposed at a Beijing conference on Friday. What should we do?

First, America should lead the way in building democratic human rights first AI standards through United Nations bodies and through the International Organization for Standardization, or ISO. America must ensure that China's authoritarian agenda does not influence global standards.

Second, we should stop American technologists from helping China build its AI surveillance state, which many have been all too eager to do. Sanctions and export controls are not enough. This Subcommittee may consider drafting a bill that metes out prison time for American executives who help develop any form of AI in partnership with a Chinese entity that could have authoritarian applications.

Third, we must strengthen our chip supply chains with our allies to ensure that China doesn't get access to critical AI logic chips. We should treat the CHIPS and Science Act as the starting point and not the last step for this goal. We can better coordinate with our partners, South Korea, Taiwan, and Japan, by upgrading the Chip 4 talks now underway into a formal R&D consortium.

As we enter the unprecedented age of generative AI, we must not allow China, a one-party authoritarian state, to infect the global ecosystem. We have seen the CCP's willingness to carry out genocide against its people with the help of AI surveillance systems. Now we must find ways to ensure that the words "never again" hold true. Thank you, Senators, for having me here today. I look forward to answering your questions.

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

Chair OSSOFF. Thank you, Mr. Cain, and thank you to all of our panelists for your opening statements. Ms. DeStefano, every parent in America can imagine the terror, the bone-chilling experience that you had, but you went through it. What was it like to hear a voice that you believed was your own daughter's—that you believed was your own daughter's expressing such distress?

Ms. DESTEFANO. It was the most horrified I've ever felt in my life, second to actually being bedside to our youngest son, who almost passed away from a rare disorder but luckily survived. It took me back to that place where you're just sitting there helpless. You don't know what to do. You don't know what to do next, where to go. The pain, and the fear, and the crying, and the sobbing, and the calling out for my help, I can't put really into words how haunting that is, and how haunting—it will probably last forever just because that's a sound you never want to hear.

Chair OSSOFF. And you were told by the authorities after reporting this crime that had you wired money or sent money as de-

manded they would investigate, but because no money was sent there would be no further investigation. Correct?

Ms. DESTEFANO. Correct. Because there——

Chair OSSOFF. And in fact——

Ms. DESTEFANO. I'm sorry, go ahead.

Chair OSSOFF. Go ahead, please.

Ms. DESTEFANO. Correct. Exactly. Since no crime had been committed there was nothing for them to pursue, or then, there was no police report that they could take.

Chair OSSOFF. In fact, my staff spoke with the Scottsdale PD, and we asked about this and were told the same thing, that because you hadn't transferred money, that there wasn't much to be done in a criminal context.

We intend to look into that more deeply at existing wire fraud statutes and other State or Federal statutes that may create a criminal claim for precisely the circumstances you raised. But I think it's clear, Senator Blackburn, Mr. Chairman, that this conduct should be criminal and severely punished. So you have my commitment to identify paths to ensuring that families are protected from what you had to go through.

Dr. Madry, you have specialized—and you're here in your personal capacity. I want to emphasize this. But I think it's important for the public to understand your credentials, a substantial track record of research and leadership at MIT. You will soon be joining the team at OpenAI. You're here in your personal capacity. Based upon your experience, help the Committee to understand other types of emergent scams, con jobs, forms of fraud that can emerge similar to what Ms. DeStefano experienced.

Dr. MADRY. Thank you. And just for clarification, I actually just recently joined OpenAI. So just to clarify this.

Chair OSSOFF. Thank you.

Dr. MADRY. So yes, what Mrs. DeStefano experienced is just the beginning, not the end for sure. As I said essentially, imagine you have this technology that can impersonate humans as long as you don't have to see them, like, in the real life. Imagine they are perfect copies. They can really deceive you. Actually, they can be better at deceiving you than many humans would be because they can pay attention to subtle cues in your speech and kind of your cognitive biases.

So now imagine that someone can just master, you know, thousands of copies of such agents, you know. What are the possibilities? There is many. I go over many in my written statement. You can field persuasion campaigns using that. You can imagine that this is how we do advertising in the future. It's actually quite scary if you start, you know, to use your imagination.

Chair OSSOFF. And in fact, just last week, Dr. Madry, the FBI issued a warning that scammers are using AI to create fake pornographic videos of victims using images and clips commonly found on their social media accounts. And as the full Judiciary Committee Chair Durbin and Senator Blackburn works through some of the legislation that we are currently moving on, child sex abuse material, this is an area that will require our study.

With my remaining time on this first round, Dr. Madry, I'd like to address at the other end of the spectrum, not the daily threats

to safety, security but what many are discussing as the emergence of potential existential threats through lowering the cost of access to technologies that enable mass destruction, like the development of bioweapons, or catastrophic cybersecurity events, or even the emergence of properties of these technologies not yet foreseeable that could place the species at risk.

How much credence do you give these warnings? Do you think they're overblown? Or do you think we need to be deeply, deeply concerned about existential risk?

Dr. MADRY. I think we should be seriously concerned because, again, some of these, in particular the ones about making it easier to build bioweapons or use them for better, you know, breaching security, they are already here. So, like, it's a theoretical possibility. So yes. So this is something we should be worrying about right now.

Chair OSSOFF. We'll get into that in more detail later. Senator Blackburn.

Senator BLACKBURN. Thank you, Mr. Chairman. Thank you all for your testimony today. Ms. DeStefano, I cannot even imagine what you went through during that period of time. But Mr. Chairman and Chairman Durbin, I think this points out why we need to look at online stalking, online harassment, and putting some of the provisions in the online space that we have in the physical space, because to be told there's nothing that can be done after you experience this—so thank you for your words.

Mr. Cain, let me come to you. Having followed what has happened in China, and I'm so interested in what you learned from the former spies who built their surveillance network, I'm grateful for the reporting that you've done on this. And what I would like to know, and you may want to do this in writing for me, which is fine, more details on the types of AI applications that are being used to surveil citizens in China. And if it's easier to put that in writing and submit it, that'll be fine. I think it would be helpful information for us.

Mr. CAIN: Yes, so I would be certainly eager to send you something in writing, Senator Blackburn. I could also go over some of that here if you have time for it.

Senator BLACKBURN. Go ahead. And I'd also like to know who—what U.S. companies are sending technology to China that they are using for this surveillance.

Mr. CAIN. Certainly. So the Ministry of State Security intelligence officers who had recently defected drew diagrams for me. I have notebooks full of these diagrams. They show where exactly the lines of power are drawn, and what they revealed to me when it came to Uyghur populations and minority populations in particular, that this was a highly centralized system.

That all cameras, which cover nearly every square inch of this region, are scooping up facial recognition data, voice recognition data. They've also gathered biometrics on pretty much everybody in the region. And this is all scooped up to the Ministry of State Security in Beijing. This is the very top of the heap. This goes straight up to Xi Jinping himself. This is not something that anyone can argue is a local project or is being done by local authorities. It is a national plan of China's.

Senator BLACKBURN. So it's a plan, it's coordinated, it's purposeful?

Mr. CAIN. Yes, it's purposeful. And what they told me and also they showed me WhatsApp messages—by the way, the Chinese Security services use WhatsApp because even they don't trust WeChat, the Chinese version, because they think they're getting spied on. So they showed me WhatsApp messages with their spy handlers in which they're being ordered to create a nationwide project and in which the plans are to expand this globally into other countries that might want these capabilities.

Senator BLACKBURN. Thank you. Ms. Givens, we don't have—you mentioned privacy in your remarks. And of course, the EU has GDPR. We have never been able to get a privacy standard on the books.

And when you look at the development of AI, and you think about things that need to be in place before we start down this road and look at different applications, whether it's defense or logistics or banking or healthcare or entertainment, like a lot of my constituents in Tennessee. Logistics, healthcare, entertainment, they're doing some good work there. But talk about the impact of not having a national consumer privacy standard. Talk about the impact that has on AI development.

Ms. GIVENS. The need for Federal privacy law in the United States is overwhelming because of the real-world harms that are happening to people now and because of the way that we're seeding global leadership on these issues.

Just to draw a couple of examples, the way that our current privacy regime exists which is a patchwork of State laws, some sector-specific laws, relies on notice and choice. This abstract idea that users can consent to their data being taken. But we know because of the way that AI uses people's data that that simply isn't the case. We are beyond a regime where users actually opt in to any of these systems about how our information is used on a daily basis.

So we have to have baseline rules of the road established in a Federal law to limit the collection, sharing, and use of people's private information. When we look at deepfake audio and video, the source material for that is people's private photos and audio recordings that have been shared and used at scale.

You can also think in the advent of generative AI how much information we share with a search bar in any given day. Now, think about the private information people are going to be sharing with a chatbot. How do we map over to make sure that those are secure environments as well and that people can have trust in these systems for them to develop?

The last example I'll give is in the use of AI when it's used to discriminate against people in employment, lending, or housing. All of that is powered by data driven inferences that a privacy law could help address. And the final thing is that some of the model privacy laws that have been introduced get at these questions of algorithmic transparency and accountability.

So putting those two things together can be incredibly powerful. So we're getting at root cause, the vast amount of private informa-

tion that is so widely available, and then also dealing very specifically with these AI use cases as well.

Senator BLACKBURN. Thank you. Thank you, Mr. Chairman.

Chair OSSOFF. Thank you, Senator Blackburn. Senator Durbin.

Chair DURBIN. Thank you, Senator Ossoff. It's good to be back in the Human Rights Subcommittee. You're doing a great job on the Subcommittee. Interesting subject, artificial intelligence. I have this hearing today and two different briefings this afternoon.

And it's not unlike that for the last several weeks, two or three different briefings a day. And for liberal arts lawyers like myself I need them all to try to understand some of the technical concepts that we are discussing, and more equally important, impact that they're going to have on the lives of Americans across the board.

We have several bills that we've considered before the Senate Judiciary Committee which go to the subject of the social media platforms, and any responsibility they have. The interesting thing is we have five bills, all bipartisan bills, Democrats and Republican sponsors, and they all passed this Judiciary Committee unanimously. Unanimously. And the premise behind them was the notion of responsibility on the part of the social media platforms as to what they're posting.

Under Section 230 for the longest time they didn't pay much attention to what was being posted. Now they're starting to pay attention. And that's led to a very active discussion within the ranks of Democrats and Republicans on the Hill about how far we should go in holding them responsible or liable for misconduct.

Ms. Reeve Givens, welcome back to the Senate Judiciary Committee. We unanimously approved the STOP CSAM bill, a bill I introduced to crack down on proliferation of child sexual abuse materials online. Your organization, for some reason, opposed the bill. One part of the bill your organization took particular issue with is a provision that pierces Section 230 of the Communications Decency Act and allows CSAM victims to sue platforms that host, store, or otherwise make this illegal content available.

We had a classic example at a hearing. A young lady at the age of 15 thought she discovered a true boyfriend on the internet, was enticed to send sexually explicit videos and photographs to this person who put them online. She's tried to contact the social media platform that posted them. They wouldn't get back to her. They wouldn't accept any responsibility. They wouldn't remove them.

She's been going through this for 20 years now. She's attempted suicide three times. She can't hold a job because this person eventually, whoever is releasing it, finds her and releases the information and the videos again to haunt her along the way.

I heard echoes of your argument against the STOP CSAM Act in a recent interview you gave to Bloomberg in discussing potential liability of a platform like I've described, when a generative AI tool causes harm. You noted that generative AI tools, and I quote you now, "do involve users engaging in expressive conduct," end of quote. I'm not sure I understand the expressive conduct of someone who's posting sexually explicit videos of a child. And I also don't understand if it would be expressive conduct when I listened to Ms. DeStefano's experience.

It seems as though a company that releases a tool that can clone a person's voice should be able to predict some of the ways the tool would be misused. And if they don't put sufficient safety measures in place, they should be held legally accountable. That to me sounds just obvious. So I'm worried about your phrase "expressive conduct" and your opposition to our bill. Would you like to explain?

Ms. GIVENS. I would, Senator. I run an organization that focuses on human rights and the impact of technology on regular people around the world. So the issues that you're raising are quite literally the hardest set of opposing tensions that we deal with. And the reason we approach these questions the way that we do is not by any means that we want to limit the ability of victims like that to seek redress.

It's how we worry about the impact of those legislation leading to platforms who have a profit motive and who act when they're scared of liability to over-police other types of conduct that are lawful and are expressive. So we worry about the downstream effects of the heavy thumb of regulation. Now, that doesn't mean by any measure that we want companies to turn a blind eye to this or to be inactive. We believe in every force of market pressure encouraging them to take those responsibilities deeply and seriously.

Chair DURBIN. What would you mean by market pressure?

Ms. GIVENS. So for example, the way that platforms now have advertisers potentially threatening to pull their ads if they don't think that they have responsible codes of conduct on their platforms, if they're not enforcing that in meaningful ways.

And my organization actively pushes those companies ourselves to be responsible and thoughtful in how they're acting, to be transparent in what they're doing, to be consistent in their approaches. But there's the additional hammer of legal liability. We worry about the long-term effects of how that changes in platforms and leads to over-takedowns of what could be expressive conduct in other settings.

Chair DURBIN. See, you talk about the heavy thumb or whatever of government. What we have now is not a heavy thumb. We have a hands-off. We stand by the sidelines and watch this poor victim, watch what happened to Ms. DeStefano. And to argue that we are somehow suppressing the market, you know, perhaps we are asking for responsibility, accountability in the market. And if you made a decision to put a car on the road that was really cheap and you were going to make money on it, unfortunately if the brakes are awful, you pay a price for that.

So the expression of the market took second place to the safety of people driving the car and those around them. So I just have to tell you, I disagree with your premise that the market is more important than the individuals who are the victims of it.

And I think that asking people to be held accountable for what they have produced and what their actions result in is as basic as justice in America. And to ignore that we are to say Section 230 or something like it should continue and stop this child sexual abuse and material online exploitation, I think it goes way too far. Please, respond.

Ms. GIVENS. Thank you, Senator. Just to be clear, I'm not worried about protecting the market in an abstract notion. I'm worried

about protecting other users who are posting lawful content, but for whom automated content filtering and some of the other provisions that companies would use if they were worried about legal liability would lead to over-removals.

So for example, when we apply these types of mandates, if companies suddenly get worried—and this has happened in the instance of the SESTA/FOSTA bill that was passed by Congress with very noble, understandable intentions to address the scourge of sex trafficking online. We also now understand analyzing those effects that sex workers have had a harder time finding online spaces to find community and express their concerns.

And that's been documented in terms of the effects. So there is absolutely no questioning the intent of Congress and the very real harms that you're trying to address. But I'm saying that there are unintended consequences for other lawful users in the ecosystem.

Chair DURBIN. So the question is whether we accept the premise that those who have these online platforms have any responsibility to police content, particularly when we're talking about child exploitation and trafficking. For God's sake, there's got to be a line we can draw that protects the marketplace but still doesn't exploit innocent people. Thank you, Mr. Chairman.

Chair OSSOFF. Thank you, Senator Durbin. Senator Blumenthal.

Senator BLUMENTHAL. Thank you. Thanks, Senator Ossoff, for having this hearing. As one of the authors of the SESTA/FOSTA bill, I happen to be very proud of it. And the consequences that you've described for the sex workers have to be addressed. But that's not a reason—to try to protect the victims of trafficking, or the victims of CSAM, or the victims of fentanyl, or the victims of a variety, of a plethora, of other evils that the tech platforms know they are enabling and propagating and empowering.

And as somebody who has written a variety of legislation and enforced it, legislation can never be wholly good. We have to accept that there will be other consequences, intended or unintended, that we need to safeguard against. But let me just come right to the point here.

We've had a number of hearings, one of them involving Sam Altman. You referred to it, Mr. Cain. In that hearing and in the subsequent hearing held in the Committee on Intellectual Property and Copyright involving another four or five witnesses, everybody agreed Section 230 does not apply to AI. Do members of this panel disagree? And if so, please, speak up.

Dr. MADRY. I do not necessarily agree or disagree. Actually, I just don't know what is the answer here. Like, it's very clear what's happening on the technical level. Now, how do we interpret it from the legal perspective? Like, that's something that is unclear to me.

Ms. GIVENS. So I also don't have a formal position on this. I think it's going to be a—this is something courts are going to have to figure out, and it's going to be a very fact-specific inquiry. I think that the arguments for 230 protections often will not apply in generative AI systems by any measure.

The goal that Section 230 is meant to promote is allowing users to create and express themselves in an online environment. And often what we're seeing with generative AI is less about user expression, right? It's a user putting in a query for medical advice,

and that's very likely just the company spewing something back as opposed to something that the user is actually generating or creating.

So I think there is—and Senator Wyden has been clear about this as well. It is a very different fact pattern than what 230 was generated for. The one exception that I think of is when an individual, for example, might use an image generating tool for their own expressive purposes. It's them that's using that tool in a particular manner. That's where I think there's just a little bit of a question of where the facts will go and we need to think that through.

Senator BLUMENTHAL. So how would you enforce Section 230? What, by deepfake? By impersonation? I'm not sure I understand.

Ms. GIVENS. Oh, no. I'm sorry, Senator. To be clear, there are different factual scenarios for how generative AI might be used and where the line of liability should fall. There are the developers of the AI tool, there are the deployers of the generative AI tool, and then there are users. And they're all making different choices that might trigger different types of liability.

Senator BLUMENTHAL. But what about the platforms?

Ms. GIVENS. So it depends what we're counting as the platform in this instant, right? So for example, a generative AI tool that would not typically fall in the bucket of 230 by any measure. So the point that I'm making is that there are moments when it is actually going to be the end user that is making that tool do something intentionally bad. And my instinct here is that the user should be the one who is mainly responsible for what they are doing.

Senator BLUMENTHAL. Well, my takeaway from this panel is that we need to clarify Section 230 to say it doesn't apply to AI. Because if it does, we're in a whole new world of hurt.

Ms. GIVENS. I do think that there's an awful lot that courts will be able to figure out just through this simple question of where is aiding and abetting liability. The simple, the straightforward shield if you can't litigate it that 230 provides, I agree that that very unlikely applies to generative AI tools at all.

Instead, I think you are allowed to pierce through and then you get into the question of who's doing the conduct where. And I think that's where there's going to be a really fruitful discussion of where you apportion that liability and the responsibility that we want for the platforms for the generative AI tools to make sure they can't be misused.

Senator BLUMENTHAL. Yes, I'm not willing to let the courts legislate. I think we have a responsibility to legislate. And we have a responsibility to protect people who may be victims, and we're moving in that direction. We're also working on legislation that would establish an oversight agency, some independent entity that would set common-sense rules, and a licensing regime for certain uses of AI, not to discourage any form of free expression either through that legislation or through any rewriting of Section 230.

We want to encourage innovation and startups in AI the same way that Google and Facebook were able to take on the IBMs of the world, the great giants, through their innovation. And we want

to support and encourage people who are doing it in their garages, startups.

But we also want to avoid repeating the mistakes that we've done through social media, which literally got halfway around the world, as Mark Twain used to say about lies, before the Congress got out of bed. And we're still trying to make up for lost time there through the Kids Online Safety Act and other measures.

Basic rules of the road can be a sustainable foundation on how we move ahead with AI. So I would be interested—my time is up here, but any of your written comments on these kinds of proposals would be greatly welcome as we go forward. Thanks, Mr. Chairman.

Chair OSSOFF. Thank you, Senator Blumenthal. And Ms. Givens, I think that this discussion about who has liability is essential as we discuss potential regulation and how issues that arise from this technology may be treated in the courts. Let's discuss that in a civil rights and criminal justice context.

As you noted in your opening statement, there was recently a man, Mr. Reid in Atlanta, Georgia, who was arrested and held in jail for 6 days on suspicion of a crime committed in a different State because of a false match through facial recognition technology. So let's just begin by acknowledging for the record, Ms. Givens, these tools and technologies are hardly foolproof. Correct?

Ms. GIVENS. That is absolutely right. And we are seeing the errors in those systems deeply impact people's lives today.

Chair OSSOFF. There are a whole range of applications in the criminal justice context that raise troubling questions. Let's focus on this facial recognition question for the moment, and let's discuss a hypothetical.

If a police department uses an AI-driven facial recognition tool and makes an arrest, or perhaps the prosecutor brings a charge on the basis of a match using that tool, and it turns out that the arrested or charged individual is innocent, and a study reveals that the underlying facial recognition tool has ingrained in it some racial bias, or is less accurate in matching Black faces than white faces, and a civil rights claim is brought against the department or against the DA's office, where might the liability rest?

Is it with the department, the prosecutor who used the tool? Is it with the producer of the tool? Is it with the AI model that the producer of the tool licensed? Is it with whoever curated the data that trained the AI model? Your take, please.

Ms. GIVENS. So sadly, that's not a hypothetical. We've seen that these systems do have statistically significant differences, particularly for people of darker skin. When we look at the few examples that we know in public record of misidentifications, those are all Black men so far that have been wrongly arrested. And that's only the tip of the iceberg because right now people don't know when it's an AI tool, when it's face recognition that's being used to just generate their arrest.

So there's a huge information asymmetry here where people don't even know that they are the subjects of these tools. And that's the case with face recognition. But also, many of the AI decisionmaking tools that we could also talk about today, whether it's in housing, lending, employment.

I do think without question, the responsibility first and foremost lies with law enforcement in the case of face recognition technology. If you are going to be making an arrest, you need to make sure that you are doing so under the Constitution on a reasonable basis, and you need to be complying with all of your constitutional obligations in that setting.

And right now, the accuracy concerns of face recognition raise that issue, but also other concerns as well with how the use of face recognition impinges on people's ability to express themselves, to move freely through society without thinking that they are being surveilled.

So the primary responsibility lies there, and it's not going to see action until Congress steps in to legislate. We're seeing some States and local governments step in to limit the use of face recognition by law enforcement. But we need Congress to act to make clear what the obligations are and to mandate, for example, that a warrant is required in those circumstances.

Chair OSSOFF. Let's take a case that is emerging and will likely emerge more frequently when we think about the predictive uses of this technology. How vast data sets, much of it foraged from public domain, or of course in the case of Federal or State or local agencies from law enforcement databases or data sets that they may purchase to which they license access, being aggregated, analyzed to train models that make predictions about risk of criminal activity geographically or even at an individual level.

Let's just take an example where such a model is trained based upon public domain and open-source information, or such predictions might be made using open-source and publicly available information. Is there some point at which that becomes itself a form of search by the state?

Ms. GIVENS. Those methods raise very deep questions as to what could amount to probable cause. The types of examples that you're talking about here come up, for instance, where law enforcement is doing social media analysis to try and indicate who might be culpable of a crime to look at those types of data points, or as you mentioned, to do inference analysis.

And all of them—both raise real questions about the accuracy and the likelihood of what they are generating really being a legitimate foundation for law enforcement action.

They also raise—their simple use raise real questions for our democracy when we look at the vast amount of data that is being collected.

Again, going back to this question of commercial data privacy practices, these are people's Facebook profiles, and the images that they've shared of themselves, and what they think of as private settings now giving rise to law enforcement uses. Law enforcement can purchase data about people from a data broker and use that for their investigation, not having to go through any of the traditional law enforcement requirements for a search.

So what we are seeing is the proliferation of data creating these mechanisms for law enforcement to be able to circumvent their legal obligations, and that's something we need to fundamentally worry about as well.

Chair OSSOFF. Let's think about it in the context of fair housing laws or laws and precedent that establish parameters for access to public facilities. Of course, technology is emerging and will be used by property owners to screen applicants for tenancy embedded within which may be racial bias, which on its face would violate fair housing laws. How are you seeing these threats to civil liberties and consumer rights emerging, and how should Congress be thinking about responding?

Ms. GIVENS. So sadly, that is also not a hypothetical. Those are harms that we're seeing right now. I can give two specific instances.

One is a growing number of landlords who are using face recognition technology, ostensibly for security purposes on their campuses. But actually, what they are doing is also being able to identify somebody who is in arrears on their rent, for example, and being able to identify them in that way instead. So this is surveillance capabilities for security being instead misused in a way that impinges on people's fundamental freedoms to go in and out of their home.

The other area, as you mentioned, is in access to housing. We also see this in access to jobs and access to credit and lending. Increasingly, we are seeing private sector tools that draw together inferences and data points about people.

For example, their education, history, whether or not they've ever had an arrest record against them, what their credit score is, whether or not they've ever been in default on something, and compiling all of those to see if somebody is suitable and eligible as a tenant or as an employee or for a particular setting of credit.

Evidence shows that those often are not good predictors, and they're not fair predictors of whether or not somebody should be able to have, you know, access to an apartment.

We know for example that education records, if you look at that, and arrest records in our country skew demographically against historically marginalized communities. And so when we're looking at that versus much more objective data, like, "Have you paid your utility bills on time the past couple of months," we're ingraining metrics and values that can really entrench and deepen inequality.

And right now, there is no oversight of this. There's no requirement to be transparent about it that's meaningfully enforced, which is why it's important that Federal agencies, the Consumer Financial Protection Bureau is doing work on this, the Equal Employment Opportunity Commission is doing work on this, Congress could also be using its oversight powers to look at the existing civil rights protections that we have, see how well they're rising to this moment, and then fill in the gaps to make sure people are really protected.

Chair OSSOFF. Thank you, Ms. Givens. Dr. Madry, would you say that the rate at which this technology is growing in capability is linear or exponential? And how do you foresee that trend developing over time?

Dr. MADRY. So definitely, if you look at the past 10 years, I would say, exponential. In a sense there are things that 10 years ago seemed like a complete science fiction to me that now are just reality. Of course, you know, it's hard to make predictions espe-

cially about the future, but again, if the last 10 years tell us anything, we should expect quite a lot of, you know, rapid developments ahead of us. But, of course, only time will tell.

Chair OSSOFF. To protect against risk, for example, of manipulation of biolabs or attacks on nuclear sites and critical infrastructure, is your view that emphasis at this time should be on guardrails embedded in the AI systems themselves, or on defensive technology and innovation in cybersecurity?

Dr. MADRY. Well, the answer should be both because essentially, like, I think the U.S. Government should really get its hands dirty and actually develop AI themselves. And that would be on the defensive part. But yes, the guardrails are definitely something to think about. We should just keep in mind that we can only put the guardrails on things that we control, so essentially things that are developed by law-abiding U.S. or other international companies. But yes, like, we should do both.

Chair OSSOFF. Things that we can control and things within our jurisdiction. Mr. Cain, you suggested in your opening remarks the need for international organizations, whether the U.N. or ISO, to be engaged to develop global standards. Talk a bit more about your vision for that and how you might see it working.

Mr. CAIN. Yes, so thank you, Senator. The ISO has already passed a number of global standards and also the UNESCO. So the United Nations Science and Education Organization has also done its own standards.

One of the problems with what's been passed so far is that they have allowed China to make these moves that sound that—as if they are public relations moves.

So in 2021, there was one standard passed out at UNESCO, and later that year the Chinese government said that it was going to drop using AI for its social credit systems in China to follow these particular standards. But I have sources in many of these Chinese firms that develop social credit, and they tell me that AI is still being used just wildly without any guardrails whatsoever. There's little that that particular international standard did.

So my vision would be something that is more enforceable under the law, something that would be required for U.N. member states to actually enforce or to create legislation, you know, within each member state. So something similar to the International Criminal Court or the European Union.

Now, you know, I do know that this is not something that could happen overnight, but with the extent of the technology that we're now dealing with I think this might be the only way to ensure that bad actors like China or even Russia or others can't, you know, trample over the international order.

Chair OSSOFF. Ms. Givens, your perspective please on international law and artificial intelligence.

Ms. GIVENS. So I think we absolutely need international cooperation. Number one, these tools are used across borders. They impact people across borders. And number two, I think the values that we bring to that conversation, to Mr. Cain's point, are deeply important, and the U.S. needs to be in these spaces.

There are areas where that's happening now, but we should think more about that, how that's integrated with the domestic

agenda. So, for example, the U.S. and the EU Trade and Technology Council is an ongoing cooperative effort between the U.S. and the European Union to have alignment as they think about the governance of AI, and in particular, to develop a shared vocabulary around how AI systems work and where regulatory interventions can fit in, and to talk about what standards for safety and mitigating some of the harms we're talking about online look like.

So I think that's a really important example of how cooperation can happen. There's another instance, though, where we need to be careful of international agreements actually undermining our efforts to regulate these spaces at home. So, for example, right now a number of advocacy organizations and Members of Congress have spoken out to warn the administration that in a trade agreement that has intellectual property protections, for example, you don't inadvertently undermine domestic efforts to demand transparency of AI systems.

So I raise that because it's an important example of how international and domestic conversations need to sync up with one another, and we need to make sure that we are able to project our vision of democratic governance and human rights in these settings around the world.

Chair OSSOFF. Thank you, Ms. Givens. Senator Blackburn.

Senator BLACKBURN. Thank you. Mr. Cain, I wanted to come back to you on the second part of my initial question to you about what technologies, what U.S. companies may be sending technology to China and the CCP that they could use. And do you know of any American companies that were involved in creating or funding AI tech that was used to surveil citizens in China?

Mr. CAIN. Yes. One of the greatest perpetrators of what you are saying is Microsoft. Microsoft has run an AI laboratory in Beijing since the late 1990s. It's called Microsoft Research Asia. This is the laboratory that went on for two decades to train many of the top AI technologists and developers in China, many of whom went on to now-sanctioned firms, such as SenseTime, Megvii and—I'm sorry, the last firm escapes me at the moment, but major, major multibillion-dollar firms.

Some of these individuals are now sanctioned in addition to their companies. And they were directly involved in creating the facial recognition and the voice recognition technologies that were sold directly to Chinese authorities, to the Defense Ministry, to the Public Security Bureau, and to the State Security Bureau. Microsoft has created itself at the core of the Chinese AI ecosystem.

And even just—I have an article here in the Financial Times just reported just this week. So Microsoft will be moving many of the AI developers from this laboratory to Vancouver because according to the article, there have been many internal discussions about the problematic nature of what has been happening over there. That they're getting tangled up in just a really bad situation and they need to separate these operations.

Senator BLACKBURN. Okay. And then you mentioned TikTok and ByteDance in your testimony. So touch on how you've witnessed the CCP use TikTok and ByteDance to help build out their surveillance state.

Mr. CAIN. Yes. ByteDance is—you know, here in America we know TikTok as the social media app with the dancing videos and the cat videos. In China, ByteDance was directly involved in working with the Ministry of Public Security to spread propaganda about the Uyghur genocide and about the atrocities against human rights there. This was a formal contract. This was set up. It was a formal relationship. It did not happen under the radar. It's something that ByteDance was directly involved in.

And, you know, personally I find it a bit ludicrous that a company that's involved in a genocide overseas can operate so openly in America. I think that's a gross, just horrific, you know, just a failure to uphold basic principles of rule of law and human rights and democracy here. And for that reason, I think TikTok should be severely restricted on U.S. soil.

Senator BLACKBURN. Okay, thank you. Thank you, Mr. Chairman.

Chair OSSOFF. Dr. Madry, in some ways there's a tension between what we've thought of traditionally in the AI space training models to recognize certain patterns and images and to make predictions and on the generative side, the production of images, video, audio.

And there's the potential for the pattern recognition capabilities of AI models to be a countermeasure against the production of counterfeit, inauthentic content such as what terrorized Ms. DeStefano. Which capability is advancing more rapidly? The ability to detect what is fake or the ability to produce it? And is that something inherent technically or does that just reflect where the R&D money is going right now?

Dr. MADRY. That's an excellent question. So in general, indeed, there is this kind of complementarity of, you know, recognizing if something is fake or not versus being able to generate something that can pass as being real or not.

And in some sense, like, the unfortunate dynamics here is that if I have a good detector of a fake content I can turn, there is a technical reason for that, it into a even better generator of bad content. So what we are essentially, like, facing here is this kind of cat-and-mouse game in which kind of we really want to be ahead on the right side.

And this brings me to the other point you mentioned, is the funding and incentives. Currently, I do not see that much incentives being provided for the detection of the deepfake. Like, as far as I know, I'm sure some of the companies are doing something, but in the research space definitely more activities on generation than on detection. Which makes sense because that's what research is about. But I would love if the Government provided, in some way, some incentives to much, much more work on the detection side.

Mr. OSSOFF. Ms. Givens.

Ms. GIVENS. I think that's absolutely right. We need extensive and quick research into deepfake detection technology and good ways to help authenticate content so that it can be trusted in how to make that as effective as possible. I do think there are also ways to strongly incentivize the companies to play their part in doing this. And a large part of that is going to be about how existing law maps onto this.

We got into a conversation about Section 230, but unlike in the 230 context, if a generative AI tool is quite literally being used to generate a falsified image, or is allowing somebody to create child exploitation material, that's the company's own tool that is doing that specific thing and surfacing that as a result. And so that's where we may well see litigation for defamation or for other things surface onto those companies themselves.

So this is an area, and I talked about this in my testimony, where I think Congress can and should pay very close attention to whether existing laws are helping address these, how the liability is falling, help shape that conversation, and use that in addition to some of the market pressure and government pressure that's being on the companies right now, to step up on some of these questions of how their tools are being used and the content that they might generate.

And I think the combination of those two things, it's not a silver bullet but that gets us at least much further than where we are now on helping to address these types of concerns.

Chair OSSOFF. Dr. Madry.

Dr. MADRY. I just wanted to add one related piece to that, is that in some sense whenever the company that is, like, whose tools, like, is providing this AI, is developing this AI, if they cooperate they can actually give us a home field advantage in this combat because they can provide some watermarking or some other capabilities to make it easier to detect that this is a fake content.

Again, this is still all proof-of-concept prototypes right now, but it would be great to have incentives as much more work in this space. But the point is that we can kind of make it a bit easier for us to detect it if we have the cooperation of the industry here.

Chair OSSOFF. Ms. Givens, what kinds of First Amendment concerns arise?

Ms. GIVENS. So as I mentioned in my opening testimony, there are very good, lawful, legitimate reasons why people might want to manipulate images. Right? There's parody, there's my kids messing around to see what images they can create on these tools for fun as an experiment. We've seen researchers, for example, transform photos of American cities to show what they would have looked like had they been subject to the extensive bombing that happened in Syria as a way of public education.

These are all good reasons why generative images and manipulated images might have useful purposes and should be treated as a form of expressive conduct. So the tricky question comes in on how we incentivize the companies to address harmful misuses of that technology and put in the safety guards that they can to address that.

For example, there are some companies already that say images of political figures running for public office simply cannot be manipulated on their platforms. The technology doesn't allow it so that they do not contribute to election related deepfakes.

There are things that companies can do, but how we create that balance between what the companies are choosing for their content policies in a way that promotes safety but also allows parity in the expressive activities that our Constitution protects and that as a

society we will want to foster, that is the challenge before us right now on how we balance those two issues.

Chair OSSOFF. Dr. Madry which emergent capabilities or capabilities that are here today most excite you?

Dr. MADRY. Excite me? That's an interesting choice of the word.

Chair OSSOFF. Or if you're not excitable, which do you believe have the greatest potential to support and promote human flourishing, human health, human well-being, and human freedom?

Dr. MADRY. Okay, so that's different because we were talking about all the bad users. So I'm not excited about any of them, but I'm definitely very excited about many of the potential outcomes. To me, the biggest vision that I have of positive vision about AI, and hopefully it's relatively close, is essentially having this personal tutor, personal kind of, like, essentially tutor who understands us, understands our learning deficiencies if we have them, understanding how we learn, and helping us learn about different issues.

So essentially you can use generative AI to kind of help you kind of look at the solutions to your problems and seeing, you know, what mistakes you are making, explaining these mistakes and so on. So we are seeing some early work on this. In particular, Khan Academy is working on such technology and I'm extremely excited about the impact it would have on the humanity if this kind of really high-quality education could be available to everyone at minimal and ideally no cost.

Chair OSSOFF. And Ms. Givens, we'll give everyone the opportunity to say what they're most potentially enthusiastic about. But I just want to—because this question on education it raises, I'm afraid to say, Dr. Madry, a question about risk.

You know, when we think about the way that we sort children, the way that standardized testing regimes function to sort young people toward careers, toward educational opportunities, the capacity to make judgments about human potential on the basis of data that to this point was not intelligible is vast.

The potential to use it for good, to provide personalized educational experiences that meet special needs is vast. But so, too, is the potential for this to constrain human freedom and to determine the choices and futures available to a human being from a very young age. Ms. Givens, how should we be thinking about regulation or best practices or standards in education?

Ms. GIVENS. The way you phrase it is so beautifully put. This is a privacy issue. My goodness. The type of interaction that we have with those systems, all of that potential—and there is so much—if that is also used to profile you, to say what learning differences you have as you're going through an experience, if we don't have strong Federal privacy regulation, anybody could get their hand on that data and the company could just bury it somewhere deep in their terms of service, and you wouldn't even know when you start using that tool.

So this is why we need rules of the road. We need rules of the road for privacy. We need rules of the road for how people can use this information and for people to be able to sue and bring a cause of action if they are being discriminated against based on this type of information, for example.

And then, of course, you mentioned the need for responsible design. So there's legal liability but even absent individuals vindicating their rights. We also need to make sure that companies are coming into this with a mindset of safety.

And that's where entities in particular working in the education space have to be committed to equity and serving the person first, making sure that what they're doing is accommodating people's needs in learning, but not triaging the top students from the bottom and leaving the bottom just to keep circulating in that ever-reinforcing pattern.

That's where questions—it's going to be hard for Congress to very specifically mandate exactly how those tools should work. But that's where general-purpose legislation like algorithmic accountability, mandating transparency, mandating risk assessments for what types of harms might result from an algorithmic system, and having companies have to disclose how they're addressing those harms, that's how policymakers and regulators would be able to understand the risks of those tools and take action against them when they're harming people.

Chair OSSOFF. Ms. DeStefano, in many ways your family's story sets the tone for this hearing. And you have opened many eyes across the Nation to the kind of horrifying risk that Americans face from the abuse and misuse of this technology. And I'm grateful to you for coming and sharing your story with us. Before we close the hearing, are there any final reflections or comments that you'd like to make?

Ms. DESTEFANO. What I experienced was horrible. It was one of the worst 4 minutes of my life. That being said, that doesn't mean that all AI is obviously evil. Listening to a lot of different areas that it can be used for good is really inspiring. We have a young son with a genetic disorder, and my daughter, Aubrey, we also spoke about, went through speech therapy for 6 years.

The advancements and accessibility that AI can help these children grow and overcome disabilities is incredible. It was very difficult for us. That's why I knew what an unknown number would often mean, a doctor's office or hospital, through personal experience. It was very difficult to be able to get her or both of them into developmental pediatrics and speech pathology, etc., to help them improve and overcome their disabilities.

So I think AI, by allowing education or accessibility to certain types of specialized medicine and specialized care, that can be really beneficial. So I don't want to speak horribly negative about AI. What happened to me with my daughter was the tragic side of AI. But in the other sense, too, there's a lot of hopeful advancements that AI will do to improve life as well, so.

Chair OSSOFF. Thank you. Ms. DeStefano. And Dr. Madry, both Mr. Cain and Ms. Givens weighed in on international law, international agreements, potential for the need for there to be an international regulatory agency. Your view on that as a scientist, engineer, and technologist, what is it that would require inspection? What are the standards, or thresholds, or capabilities that such an entity would regulate?

Dr. MADRY. Well, essentially, usually—first of all, I think we will only be learning what it is that we should be looking for. So that's

where you want to have this structure and agency in place that has close touch and is paying attention to how things develop. If you ask me about the capability threshold, I would put it essentially roughly at the state-of-the-art right now. And then as we see how technology develops which again we could be able to keep close track of, and what are the new risks, we might either lower it or make it higher.

But yes, I would just want to understand exactly how is this AI used, for which purposes, to what extent can we mitigate certain bad users of this, and essentially also understand where we as the whole world, not only the U.S., are in terms of, you know, emerging AI capabilities. So if there is some threshold to be exceeded, well, we want to know it sooner than later.

Chair OSSOFF. Ms. Givens, what actions must Congress take to stay on the critical path toward ensuring that the emergence of this technology facilitates human flourishing and human freedom rather than enabling the abuse of power?

Ms. GIVENS. Congress needs to look at specific use cases, like the face recognition example that I gave, which probably requires specific legislation to address those harms. But then there's an across-the-board effort that Congress could make as well, which is to get to this question of mandating transparency and mandating disclosures of how companies are looking at questions of safety, validity, their fitness for purpose, whether they discriminate, whether they violate people's privacy.

We need to establish that as the baseline analysis for any company whose tool could have a high-risk use to go through that process, and to do it not just internally but to publicly disclose how they're thinking about those risks and what they are doing to mitigate those risks. We can't have accountability without that baseline rule of the road because we literally don't know how the harms are going to manifest, and we can't just have individuals trying to fight this David versus Goliath battle.

So if we talk about algorithmic accountability, Congress can step in there in a meaningful way to try and really start that conversation, and then have ongoing oversight of how well our civil rights laws and product liability laws are rising to the occasion as well. So I think there's steps Congress can take now, like legislating around algorithmic accountability, and then there's oversight power that Congress can have, too, of how the sector continues to evolve.

And above all, I think one of the big pieces—somebody mentioned earlier that they're not a technologist. Senator Durbin said that. We need non-technologists to feel they have a seat at the table. We need public voices to have a seat in these conversations. So right now, governments around the world are talking to some of the largest companies about the safety standards they're going to adopt, and that's good.

But there's a role for Congress to help make that a much more public conversation, where civil society advocates and regular people have a seat at the table as well. And that's another area where Congress can use its oversight authorities now to help drive that conversation forward quickly but in a meaningful way.

Chair OSSOFF. I want to thank all of our witnesses for appearing today and for helping us work through these questions. I thank my colleagues who attended for a productive discussion.

After what we've heard today about the risks and the opportunities, it is clear that the Senate must continue and accelerate our study of machine learning, of artificial intelligence, and Ms. Givens, to the point you made and Senator Blackburn made, get our act together on a national privacy law. Without national privacy legislation, our efforts to control the abuse of these technologies are substantially reduced. And so that is an urgent task for the U.S. Congress.

The hearing record will remain open for 1 week for statements to be submitted into the record. Questions for the record may be submitted by Senators by 5 p.m. on Tuesday, June 21st. The hearing is adjourned.

[Whereupon, at 4:11 p.m., the hearing was adjourned.]

[Additional material submitted for the record follows.]

APPENDIX

Additional Material Submitted for the Record

Witness List
Hearing before the
Senate Committee on the Judiciary
Subcommittee on Human Rights and the Law

"Artificial Intelligence and Human Rights"

Tuesday, June 13, 2023
Dirksen Senate Office Building, Room 226
2:30 p.m.

Jennifer DeStefano
Victim of AI Deepfake Kidnapping / Extortion Scam
Scottsdale, AZ

Professor Aleksander Mądry
Cadence Design Systems Professor of Computing
Massachusetts Institute of Technology
Cambridge, MA

Alexandra Reeve Givens
CEO
Center for Democracy & Technology
Washington, D.C.

Geoffrey Cain
Senior Fellow
Foundation for American Innovation
Chicago, IL

FOUNDATION FOR AMERICAN INNOVATION

June 13, 2023

Geoffrey Cain
Senior Fellow, Foundation for American Innovation
Written Testimony for U.S. Senate Committee on the Judiciary
Subcommittee on Human Rights and the Law

"America, the Vanguard of Democracy, Must Stand Up to China's AI Totalitarianism"

Chairman Ossoff, Ranking Member Blackburn, and members of the Subcommittee:

Thank you for the opportunity to testify today. My testimony has two purposes:

1. First, to outline how China has created the world's most sophisticated and terrifying surveillance state using novel artificial intelligence (AI) technologies, and how American business elites helped make this happen.

2. Second, to suggest ways that the US can defend the use of AI with respect to democracy and human rights, to ensure that the Chinese Communist Party (CCP) cannot advance its malign global agenda with AI tools.

With AI, America's elites have learned little about the perils of engaging with China's one-party authoritarian state

On Friday, OpenAI CEO Sam Altman dialed into the annual conference at the Beijing Academy of Artificial Intelligence, three weeks after he testified before another subcommittee here at the Senate Judiciary Committee. He called on the People's Republic of China—a one-party authoritarian state that has used AI to carry out genocide against an ethnic minority—to help shape global AI safety guardrails. "With the emergence of increasingly powerful AI systems," he said, "the stakes for global cooperation have never been higher."[1]

To anyone who's lived in China, this was a curious and mind-boggling call to action. The Chinese Communist Party (CCP) has engineered a vast AI-powered surveillance system literally called "Sky Net." It runs AI-powered "alarms" that notify the police and intelligence services when someone unfurls a banner,[2] when a foreign journalist is traveling to certain parts of the country,[3] and when someone from an ethnic minority is

---

[1] Sarah Zheng, "OpenAI's CEO Calls on China to Help Shape AI Safety Guidelines," Bloomberg Technology, June 9, 2023, https://www.bloomberg.com/news/articles/2023-06-10/openai-s-ceo-altman-calls-on-china-to-help-shape-ai-safety-guidelines.
[2] Gulchehra Hoja, "In China, AI Cameras alert police when a banner is unfurled," *Radio Free Asia*, June 5, 2023, https://www.rfa.org/english/news/china/surveillance-06052023142155.html.
[3] Jimmy Quinn, "'Total Security State': Shanghai Intensifies Surveillance of Foreign Journalists Who Go to Xinjiang," *National Review*, May 2, 2023, https://www.nationalreview.com/corner/total-security-state-shanghai-intensifies-surveillance-of-foreign-journalists-who-go-to-xinjiang/.

present.[4] The government accuses entire groups, such as Muslim Uyghurs, of posing a terrorist threat, and relentlessly persecutes them with the use of AI tools.

It sounds like a dystopian science fiction story—think *1984* or *Minority Report*—but the CCP's AI totalitarianism has become a fact of daily life for the more than 1.4 billion people in China. In fact, the Chinese technologists who spoke at the same conference as Mr. Altman were some of the very people who built this monstrosity. They were executives at iFlyTek and Huawei, two AI giants and are heavily sanctioned by the US government for their involvement in human rights abuses.[5] If Mr. Altman plans on cooperating with China's AI developers, he better figure out who he's working with.

I've witnessed the results of their work firsthand. As an investigative journalist formerly in China, I was among the first people to document and expose the horrific surveillance state that oppressed the Uyghur population in the far western region of Xinjiang. Since 2017, the atrocity has morphed into the largest internment of ethnic minorities since the Holocaust, which the US State Department calls a genocide.[6]

Chinese authorities have hauled away 1.8 million people to concentration camps—about one-tenth of the ethnic minority population in Xinjiang—and have forced many of them into slave labor.[7] Because they have read too many books or have been caught praying, they have been declared enemies of the state, despite not being formally charged with any crime. This was all with the help of the AI surveillance system that scooped up data from facial recognition, voice recognition, and a network of police cameras covering every possible square inch of the region. Party authorities told Uyghurs they wanted to "cleanse" their minds of what they called "ideological viruses."

In December 2017, I was kicked out of China while researching my book, *The Perfect Police State: An Undercover Odyssey into China's Terrifying Surveillance Dystopia of the Future.* Ever since then, the AI-fueled police state has expanded to alarming levels. In 2018, I moved to Turkey and, for three years, tracked down former intelligence officers from China's Ministry of State Security, the powerful and secretive intelligence body. They had helped set up the AI surveillance systems in Xinjiang, were targeted by those same systems because they were Uyghurs, and then defected to safety.

---

[4] Drew Harwell and Eva Dou, "Huawei tested AI software that could recognize Uighur minorities and alert police, report says," *Washington Post,* December 8, 2020, https://www.washingtonpost.com/technology/2020/12/08/huawei-tested-ai-software-that-could-recognize-uighur-minorities-alert-police-report-says/.

[5] Karen Hao, "Open AI CEO Calls for Collaboration with China to Counter AI Risks," *The Wall Street Journal,* June 10, 2023, https://www.wsj.com/articles/openai-ceo-calls-for-collaboration-with-china-to-counter-ai-risks-eda903fe.

[6] Secretary of State Michael R. Pompeo, "Determination of the Secretary of State on Atrocities in Xinjiang," U.S. Department of State, January 19, 2021, https://2017-2021.state.gov/determination-of-the-secretary-of-state-on-atrocities-in-xinjiang/index.html.

[7] Adrian Zenz, "China's Own Documents Show Potentially Genocidal Sterilization Plans in Xinjiang," *Foreign Policy,* July 1, 2020, https://www.wsj.com/articles/openai-ceo-calls-for-collaboration-with-china-to-counter-ai-risks-eda903fe.

These intelligence officers drew detailed diagrams in my possession that showed the workings of these surveillance systems and how facial recognition and voice recognition technologies helped fuel them. What they revealed was alarming, but not surprising. The highest echelons of CCP leadership held centralized control over many AI surveillance systems, as well as direct lines of influence over Chinese mega-companies such as Huawei and ByteDance. With the help of these companies, China's government had been making a concerted, malicious effort to expand these surveillance capabilities all over the world.

The development of AI is at the heart of China's global ambitions

The surveillance state that began in Xinjiang was a taste of the horrific power of AI when placed in the wrong hands. "Advanced technology is the sharp weapon of the modern state," China's President Xi Jinping said in a 2013 speech.[8] In July 2017, China unveiled its National AI Development Plan, calling AI a "historic opportunity" and pledging to align developments in AI with the government's authoritarian values. China has declared its goal as becoming the world leader in AI by 2030.[9] The goal reflects the totalitarian ambitions of President Xi, who has led the efforts to clamp down Uyghurs, Tibetans, Mongolians, and religious and political dissidents of all stripes.

Since then, we've seen the expansion of China's technology companies, using AI and other novel developments, all over the world. Huawei, the heavily sanctioned telecommunications firm, has led efforts to establish global surveillance systems, usually under the guise of AI-powered "smart cities" designed to fight crime and regulate traffic, but that in reality have been used to equip governments with the tools to spy on political dissidents. In October 2022, the FBI arrested two Chinese nationals who stood accused of bribing an undercover FBI officer to obtain inside intelligence about an investigation into Huawei.[10]

Meanwhile, ByteDance, the $220 billion mega-firm that owns TikTok, stands accused by a whistleblower of running an in-house CCP Committee that had access to all the app's data, including data stored in the US, according to a court filing.[11] Other sanctioned, lesser-known firms, such as AI facial and voice recognition companies iFlyTek, SenseTime, and Megvii, have emerged as global billion-dollar unicorns with the backing of the Chinese state and the involvement of US venture capital funds.

---

[8] Chris Buckley and Paul Mozur, "What Keeps Xi Jinping Awake at Night," *New York Times*, May 11, 2018, https://www.nytimes.com/2018/05/11/world/asia/xi-jinping-china-national-security.html.
[9] Graham Webster, Roger Creemers, Elsa Kania, and Paul Triolo, "Full Translation: China's 'New Generation Artificial Intelligence Plan,'" August 1, 2017, https://digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/
[10] Glenn Thrush and David McCabe, "Justice Dept. Charges 2 Chinese Citizens With Spying for Huawei," *New York Times*, October 24, 2022, https://www.nytimes.com/2022/10/24/us/politics/justice-dept-huawei.html.
[11] Thomas Fuller and Sapna Maheshwari, "Ex-ByteDance Executive Accuses Company of 'Lawlessness,'" *New York Times*, May 12, 2023, https://www.nytimes.com/2023/05/12/technology/tiktok-bytedance-lawsuit-china.html.

This situation is proving hard to continue in the age of technological decoupling. This month, Sequoia Capital, the preeminent venture capital firm that originally invested in Apple and Facebook, announced that it was splitting off its Chinese arm into a separate company.[12] Sequoia's China business was core to helping build China's AI industry, with a reported $22 billion stake in ByteDance, to name one of many examples.[13] Sequoia's spin-off suggests that American business executives are waking up to the unavoidable risks of doing business in China—of inadvertently helping build China's AI systems that damage human rights and the public good.

Generative AI is a threat to CCP censorship

In April 2023, the Cyberspace Administration of China announced draft regulations for generative AI, setting down potential rules that chatbot-produced content follow "socialist core values" and avoid information that undermines "state unity."[14] The CCP's goal is a continuation of its past strategy to align new technologies and censor information in line with its political values. ChatGPT has not made its service available in China, but there is already significant demand. The black market is already flourishing with offerings of overseas ChatGPT access to people in China, but these days could be numbered.[15]

Generative AI, however, is a departure from the surveillance technologies that have defined the evolution of China's political censorship. Generative AI services have the potential to empower regular people who want to produce large amounts of content that challenge government propaganda and narratives. The question is whether China's "Great Firewall"—the harsh internet censorship system—can stand up to the potential of generative AI. Will China one day see an information renaissance, with stories of the Tiananmen Square massacre and Hong Kong protestors spread across the internet through uncontrollable chatbots?

Given the CCP's enormous success at censorship so far, I believe that it will once again succeed in coercing and coopting Chinese technology firms and transforming generative AI into a tool of state oppression. American technologists will unwittingly assist CCP goals if they cooperate too eagerly with state-connected Chinese companies, institutes, and people. As we have learned over the last decade, this is the sad truth of being a technologist in China.

---

[12] Shawn Johnson, "Neil Shen goes it alone in China after Sequoia split," *Financial Times*, June 9, 2023, https://www.ft.com/content/179eb51a-70eb-4c79-9e74-befd0f5a02b7.
[13] Alex Kondrad, "For Top VCs, ByteDance's Historic Windfall Remains a $220 Billion Mirage," *Forbes*, May 4, 2023, https://www.forbes.com/sites/alexkonrad/2023/05/04/bytedance-scrutiny-leaves-midas-investors-waiting-billions/.
[14] Chang Che, "China Says Chatbots Must Toe the Party Line," *New York Times*, April 24, 2023, https://www.nytimes.com/2023/04/24/world/asia/china-chatbots-ai.html.
[15] Caiwei Chen, "China's ChatGPT Black Market Is Thriving," *Wired*, March 7, 2023, https://www.wired.com/story/chinas-chatgpt-black-market-baidu/.

The US must use its global technological leadership to protect democracy and human rights from China's AI threats

The CCP is the greatest threat to human rights and democracy around the world. Although China is quickly catching up to US innovation, the US remains the leader in AI development. We must abandon the misguided idealism of working with Chinese companies and government bodies with the hope that AI will change the political system, allow for the opening of democratic discourse, and create safer global AI regulations. Rather than helping advance innovation, we will be doing the world a disservice by handing the keys to the CCP. Under Chinese law, these advanced AI applications will inevitably be used to oppress human rights and expand China's authoritarian footprint.

Rather, we should use our position of strength and our democratic values to carry out a two-fold strategy. First, AI talent and innovation must flow towards the direction of America and its allies. We must influence global AI standards, attract global AI talent away from China, and secure our software and hardware ecosystems from China's malign influences. Second, the most advanced American technologies and investments must not be allowed to flow in the direction of China. We must work against China's ambitions to develop advanced AI systems, influence global standards, and oppress dissidents around the world. The specific policy steps are as follows:

1. **The US must take the lead in developing global AI standards that uphold human rights and democratic values.**

   The CCP has loudly used multilateral membership bodies—the United Nations, the World Health Organization, and so forth—to shape global technology and science standards in its interests and to make countries all over the world dependent on Chinese technological innovation. The US must not shirk its global leadership, which would mean ceding ground to China and abandoning our allies in a moment of global struggle.

   In November 2021, 193 countries adopted the first-ever global agreement of AI ethics under the United Nations Educational, Scientific, and Cultural Organization (UNESCO), calling for a "do no harm" principle, personal data protection, and measures to prevent fairness and non-discrimination.[16] The US should leverage other United Nations bodies and the International Organization for Standardization (ISO) to build democratic AI principles and ensure that China's authoritarian goals do not crush the principles of human rights.

2. **American companies that help build China's oppressive AI ecosystem must be held accountable.**

[16] UNESCO, "Recommendations on the ethics of artificial intelligence," November 2021, https://www.unesco.org/en/articles/unesco-adopts-first-global-standard-ethics-artificial-intelligence.

China built its AI surveillance apparatus with the connivance and complacency of major American technology firms. The science corporation Thermo Fisher, for example, was caught selling DNA collection equipment directly to Xinjiang police authorities who used them for mass gathering of genetic data on the minority Uyghur population.[17] Since the late 1990s, Microsoft has established itself as the training ground for China's AI elites through its Beijing-based laboratory, Microsoft Research Asia. The laboratory has trained many of the AI leaders and developers who went on to found or join the executive leadership of rights-abusing firms such as Sensetime, Megvii, and iFlyTek. Beginning in 2019, the US government has sanctioned these individuals and their companies.[18]

So far, American technology giants have faced no punishment for their involvement in China's surveillance state. This subcommittee may consider drafting a bill that requires public corporations to publish their due diligence reports on their activities in China and the risks they have encountered with regards to human rights there. The subcommittee may also consider drafting a bill that criminalizes specific American business activities in China that are likely to support, directly or indirectly, human rights abuses by the CCP. This would include prison time for American business executives involved helping develop any form of AI in partnership with a Chinese entity, if the CCP will likely use that technology for the oppression of human rights and democratic values.

3. **Because Chinese software companies are required to partake in Chinese state intelligence operations, they should be compelled to separate their American businesses.**

Over the past decade, China has enacted a raft of draconian laws, such as the National Security Law and the National Intelligence Law, that require people in China to assist the government in intelligence-gathering when called upon, among other requirements.[19] While we in America have a system of due process and checks and balances that can guard against data overreach, in China no such rights exist. The private and personal data of Americans is not safe in the hands of Chinese-owned apps such as TikTok and Temu, whose owners and employees in China are required to hand over data to the state if it's requested.

Apps like TikTok are beginning to form the core of the US information environment, with sophisticated algorithms that recommend highly addictive

[17] Human Rights Watch, "China: Minority Region Collects DNA from Millions," December 13, 2017, https://www.hrw.org/news/2017/12/13/china-minority-region-collects-dna-millions.
[18] Kate Kaye, "Microsoft helped build AI in China. Chinese AI helped build Microsoft," *Protocol*, November 2, 2022, https://www.protocol.com/enterprise/us-china-ai-microsoft-research.
[19] Bonnie Girard, "The Real Danger of China's National Intelligence Law," February 23, 2019, https://thediplomat.com/2019/02/the-real-danger-of-chinas-national-intelligence-law/.

content, while being used to spy on US citizens.[20] This is a gaping breach of our ability to protect democratic values and human rights here in the US. In the event of conflict with China—an increasing likelihood with China's aggressive military posture—these apps have the potential to become misinformation machines designed to manipulate Americans with sophisticated and algorithmic propaganda. The solution is to force these firms to spin off their American operations into separate companies, ensuring their safety from CPP meddling.

4. **America and its allies must secure and coordinate global supply chains for advanced AI logic chips.**

   The US has made remarkable progress in legislating and implementing export controls that prevent American firms from selling advanced chips and their components to China. In October 2022, the Biden administration implemented the most recent round of sanctions, restricting the export of certain services and equipment to China, effectively placing China generations behind American chip technologies for the latest AI applications.[21] Four months later, in February 2023, the Department of Commerce opened the first round of company grants under the CHIPS and Science Act, hoping to reshore semiconductor manufacturing capabilities and make the US more self-sufficient.[22]

   The CHIPS and Science Act, however, is the starting point and not the last step. Advanced semiconductors are the most complex devices that humankind has ever made—and they cannot simply be manufactured end-to-end in the US. Chip supply chains depend on thousands of suppliers all over the world. The US needs to better coordinate with its key chip-producing and component-producing partners—South Korea, Taiwan, Japan, and the Netherlands—by upgrading the "Chip 4" talks into a formal consortium for coordinating R&D innovations.

   The upgrade will enhance the implementation of the CHIPS and Science Act and the future of AI technologies by adding an element of multilateralism. Our technological partners will have better reason to believe their contributions to the US manufacturing ecosystem are profitable and worthwhile, a hedge against CCP aggression. If we can form a true semiconductor alliance, China will be unable to bully individual countries into supplying critical chip technologies for its AI systems.

---

[20] Emily Baker-White, "TikTok Spied on Forbes Journalists," *Forbes*, December 2, 2022, https://www.forbes.com/sites/emilybaker-white/2022/12/22/tiktok-tracks-forbes-journalists-bytedance/?sh=3b7173b97da5.
[21] Demetri Sevastopulo and Kathrin Hille, "US hits China with sweeping tech export controls," *Financial Times*, October 7, 2022, https://www.ft.com/content/6825bee4-52a7-4c86-b1aa-31c100708c3e.
[22] U.S. Department of Commerce, "Biden-Harris Administration Launches First CHIPS for America Funding Opportunity," February 28, 2023, https://www.commerce.gov/news/press-releases/2023/02/biden-harris-administration-launches-first-chips-america-funding.

As we enter the unprecedented age of generative AI, we must not allow China, a one-party authoritarian state, to infect the global AI ecosystem where it will oppress human dignity, civil liberties, and rule of law. We have seen the CCP's willingness to carry out genocide against its people with the help of AI surveillance systems. Now we must find ways to ensure that the words "never again" hold true. Thank you, Senators, for having me here today. I look forward to answering your questions.

United States Senate
Written Statement of Jennifer DeStefano
Abuses of Artificial Intelligence
June 13, 2023

Good Afternoon Senators, it is my great honor to speak with you today and to share my experience of how artificial intelligence is being weaponized to not only invoke fear and terror in the American public, but in the global community at large as it capitalizes on and redefines what we have known to be as "familiar". I would like to take this moment to thank Senator Ossoff for inviting me to be here today. I would also like to thank Senator Blackburn for your concern on this ever evolving topic and community threat. AI is revolutionizing and unraveling the very foundation of our social fabric by creating doubt and fear in what was once never questioned, the sound of a loved one's voice.

What is "familiar"? How many times have you received a phone call from your child and asked them to verify who is calling? How many times has a loved one reached out to you in despair and you stopped them to validate their identity? Did you hang up on them? Did you require to call them back to make sure you are speaking to the correct person? The answer is more than likely, never. Never have you stopped your loved one and questioned if the voice you are speaking with is really them. The sound of a loved one's voice is often never questioned. It is designed by nature, it is designed by God, as a unique identity, as unique as a fingerprint. This familiar identity is how a mother knows if it's her child crying in a room and it is how a newborn child instantly recognizes their mother.

It was a typical Friday afternoon for our family kicking off a weekend of races and rehearsals that often divide our family across the state. As the parents of four children close in age, we tend to have to "divide and conquer". My husband was with our older daughter Brie and our youngest son in Northern Arizona training for ski races. I was with our older son and youngest daughter Aubrey in the valley as she had rehearsal. Ski racing is a high risk sport and Brie had not raced in years. At age 15, she promised me she would take it easy and not hurt herself by pushing to hard. When I first received a call from an "unknown" number upon exiting my car, I was going to ignore it. On the final ring I chose to answer as "unknown" calls can often be a doctor or a hospital. I answered the phone " Hello", on the other end was our daughter Briana sobbing and crying saying "mom". At first I thought nothing of it, she had run into race gates and bruised herself before, not to worry. I casually asked her what happened as I had her on speaker walking through the parking lot to meet her sister. Briana continued with "mom, I messed up" with more crying and sobbing. Not thinking twice, I asked her again, "ok what happened?" Suddenly a man's voice barked at her to "lay down and put your head back". At that moment I started to panic. My concern escalated and I demanded to know what was going on, but nothing could have prepared me for her response. "MOM THESE BAD MEN HAVE ME, HELP ME, HELP ME!!" She begged and pleaded as the phone was taken from her. A threatening and vulgar man took over the call "Listen here, I have your daughter, you tell anyone, you call the cops, I am going to pump her stomach so full of drugs, I am going to have my way with her, drop

her in mexico and you'll never see her again!" all the while Briana was in the background desperately pleading "mom help me!!!"

With my shaking hand on the door handle to the studio, I put the man on mute, flung open the door and started screaming for help. The next few minutes were a parent's worst nightmare. I was fortunate to have a few moms at the studio who surrounded me, hearing all of the vulgar threats the man was making. One mom ran outside and called 911. Our 13 year old daughter Aubrey stood paralyzed in fear. I needed her help, her sister was in trouble and we had to find her. Another mom ran to her to aid as they started making calls to her dad, her brothers, anyone that could help us figure out what happened to Brie. The kidnapper demanded a million dollars. That was not possible and so the kidnapper decided on $50,000, in cash. At this moment, the mom who called 911 came inside and shared with me that 911 was familiar with an AI scam where they can replicate your loved one's voice. I didn't believe this was a scam. It wasn't just Brie's voice, it was her cries, it was her sobs that were unique to her. It wasn't possible to fake that I protested. She told me that AI can also replicate inflection and emotion. That gave me a little hope but still was not enough. I proceeded with the negotiations. I asked for wiring instructions and routing numbers for the $50,000 but was refused. "Oh no" the man demanded, "that's traceable, that's not how this is going to go down. We are going to come pick you up!" "What?" I shouted, "You will agree to being picked up in a white van, with a bag over your head so you don't know where we are taking you. You better have all $50k in cash otherwise both you and your daughter are dead! If you don't agree to this, you will never see your daughter again!" he screamed. I had to stall, I asked the mom on the call with 911 to send police, I needed

to stall until I had police with me. Then the mom who was making calls with Aubrey was able to get my husband on the phone. He frantically located Brie resting safely in bed. Brie had no idea what was happening. As I was negotiating the arrangements of the abduction of myself to save my daughter, the mom came to me and told me she found Brie and that she was safe. I didn't believe her. How could she be safe with her father and yet be in the possession of kidnappers? It was not making any sense. I had to speak to Brie. I could not believe she was safe until I heard her voice say she was. I asked her over and over again if it was really her, if she was really safe, again, is this really Brie, are you sure you are really safe?! My mind was whirling. I do not remember how many times I needed reassurance, but when I finally took hold of the fact she was safe, I was furious. I lashed at the men for such a horrible attempt to scam and extort money. To go so far as to fake my daughter's kidnapping was beyond the lowest of the low for money. They continued to threaten to kill Brie. I made a promise that I was going to stop them, that not only were they never going to hurt my daughter, but that they were not going to continue to harm others with their scheme. After I hung up, I collapsed to the floor in tears of relief. When I called the police to pursue the matter, unfortunately I was met with this is a prank call. That it happens often and that I am probably not in harm's way (although not a guarantee). I was offered to have a police officer call me from another "unknown" number if it would make me feel better as law enforcement numbers are also blocked. That certainly did not make me feel better. Bottom line was no actual crime had been committed, no one was physically kidnapped, and no money was transferred, period, the end.

But that wasn't the end, it couldn't be the end. If it was the end, then this nightmare would never stop. I stayed up all night paralyzed in fear. Do they know where I am? Do they know where my daughter is? How did they get her voice? How did they get her crying, her sobs that are unique to her. She is not a very public person. Are we being cyber stalked? Targeted? So many questions that I could not leave unanswered, so I turned to our community and the response was overwhelming! Friends and neighbors came out of the woodwork with their stories. Kidnapping phone calls coming from their children's phones, bags of money being driven halfway to Mexico, even voices of young children nowhere to be found on social media and who do not have phones, the stories kept pouring in. Even my own mother received a call with my brother's voice claiming to be in an accident and needing money for the hospital bill! My mother is hard of hearing and quite spunky. After having the caller repeat the request multiple times, she realized the language used was not something my brother would say. She told the caller to call their real mother and hung up. The common response the victims received from authorities was that nothing could be done. In fact, one mother I know personally shared with me how she was even mocked by her son's school and security officer. She called his school frantically trying to locate her son when she received a call from him that he had been kidnapped. He even used his unique nickname during the call to self identify. Fortunately he was safe in class and she was told "this happens all the time" as her fear was dismissed. "It's the most frustrating, maddening, scary and invaded I've felt...my fear is that it is only a matter of time until someone actually follows through with the threat", she told me as she has been living in fear and concern for her son's safety ever sense.

Money scams have been around for thousands of years. We have all heard of "snake oil" and remember the days of "swap land" sold as paradise in Florida. This is entirely

different. This is terrorizing with lasting post traumatic stress. Even months later, sharing the story shakes me to my core. It was my daughter's voice. It was her cries, her sobs. It was the way she spoke. I will never be able to shake that voice out of mind. It's every parents' worst nightmare to hear your child pleading with fear and pain, knowing that they are being harmed and you are helpless and desperate. The longer this form of terror remains unpunishable, the farther and more egregious it will become. The thought crossed my mind before I hung on the "kidnappers" to follow through with the physical abduction of me. Was that what would it take to bring an end to this? Was that what it would take in order to have a pursuable criminal offense?

As our world moves at a lightning fast pace, the human element of familiarity that lays foundation to our social fabric of what is "known" and what is "truth", is being revolutionized with Artificial Intelligence. Some for good, and some for evil. No longer can we trust "seeing is believing", "I heard it with my own ears" nor even the sound of our own child's voice. This concept redefines and rewrites what the very meaning of "familiarity" means. Familiarity is defined as "the quality of being well known or knowledge of something" and further is defined as "relaxed friendliness or intimacy between people." Familiar and family share the root word "Famil" which establishes strength of a relationship between one person and another. I ask you, when your mother calls, are you going to hang up and call her back to make sure it is really her? When your child calls you in need of help, will you disconnect the call and say I don't believe its really you? Is this our new norm? Is this the future we are creating by enabling this abuse of Artificial Intelligence without consequence?

I want to thank you for your time and attention today. Congress has a large and looming task ahead. How do we move forward as a community with this haunting reality that is plaguing us? If left uncontrolled, unguarded and without consequence, it will rewrite our understanding and perception what is and what is not truth. It will erode our sense of "familiar" as it corrodes our confidence in what is real and what is not. This is a non-partisan matter and I have seen the hands reach across the aisle in unified concern. That gives me great hope. How to contain the ever evolving Artificial Intelligence and its unknowns, is not an easy task. My sincere thanks and humble appreciation for your time and attention today. I thank all of you, and especially Senator Ossoff and Congress at large, for tirelessly taking action to keep our community and world safe from the hands of evil. I am one person, one story, but I am not the only one and I certainly will not be the last one unless action is taken. I wish you God's speed.

Testimony of Alexandra Reeve Givens
President & CEO, Center for Democracy & Technology

For the U.S. Senate Committee on the Judiciary Subcommittee on Human Rights and the Law
Hearing Entitled "Artificial Intelligence and Human Rights"

June 13, 2023

Chair Ossoff, Ranking Member Blackburn and other members of the Subcommittee, thank you for inviting me to testify today on the important issue of AI and human rights. The world's attention is rightly focused on the possibilities and the risks of AI systems. As policymakers look to address potential harms and promote responsible innovation, it is essential that they do so with a focus on human rights – and in particular, with the conviction that fundamental rights and freedoms belong inalienably to all people, including the rights to liberty, privacy, freedom of expression and opinion, peaceful assembly, and equal treatment before the law.[1]

AI systems are already being used in ways that threaten these rights, and rapid advancements in generative AI and text and image analysis will exacerbate the risks. Today I will focus on two distinct areas where AI harms are already being felt: the use of face recognition and biometric surveillance capabilities by law enforcement, and the impact of generative AI on elections and democratic discourse. For reasons I will explain in my testimony, these applications of AI are vastly different from one another, with different considerations at stake as Congress considers appropriate policy interventions.

Of course, these areas are not the only ways in which AI is impacting human rights. In previous testimony before the U.S. Senate Committee on Homeland Security and Government Affairs, I described risks posed by AI systems to people's civil rights and access to economic opportunities – for example when people are applying for jobs, housing, or credit – and potential policy responses.[2] I also described how AI is being used in ways that jeopardize the fair administration of public benefits programs, and steps the government should take to protect people's access to basic services and due process rights. Those issues are ripe and important priorities for government intervention.

At a time when many are discussing the long term existential risks of AI systems, there are concrete issues on which Congress and the U.S. government can act *today* — and, in doing so, demonstrate what it means to ensure AI is developed in a manner that centers democratic values and human rights.

[1] United Nations General Assembly. The Universal Declaration of Human Rights (UDHR). New York: United Nations General Assembly, 1948.
[2] Alexandra Reeve Givens, "Press Release: In Senate Testimony, CDT CEO Alexandra Givens Calls For Cross-Society Effort in Addressing Risks of AI", Center for Democracy & Technology, March 8, 2023, https://cdt.org/insights/press-release-in-senate-testimony-cdt-ceo-alexandra-givens-calls-for-cross-society-effort-in-addressing-risks-of-ai/.

AI & Government Surveillance

Last fall, many of us were inspired by the images of brave Iranian women protesting the death of 22-year-old Mahsa Amini after she was arrested for allegedly improperly wearing the hijab. But we were not the only ones watching those protests. In Iran today, face recognition technology allows the government to identify protestors and take action against them. Demonstrators have received text messages from local police stating that they were observed at a protest and should not join further demonstrations.[3] Iranian officials also announced that they would use face recognition in public spaces to detect and identify women who were not "correctly" wearing a hijab.[4] A member of parliament explained that women who dress improperly would receive text message warnings, followed by penalties such as their bank accounts being blocked. In Iran, citizens must use biometric national identity cards to receive pensions and food rations, open bank accounts and access the domestic internet – making these threats of automated punishments all too real. In this context, AI systems are enabling a repressive regime to identify dissenters, subject them to pervasive surveillance, and then automate their punishment.

Face recognition technology has been used in similar ways by the Chinese government, to promote social control through mass enforcement and public shaming of minor offenses such as jaywalking,[5] as well as for its notorious treatment of China's Uyghur minority.[6] Face recognition has also been used to identify protestors in Russia, Hong Kong and Uganda, among other countries.[7]

Such examples may feel far from the United States, but the technical capabilities exist here, and we do not have adequate legal frameworks to address them. In the U.S. there have already been abuses: In 2020, police in multiple Florida cities used facial recognition to identify and catalog activists engaging in peaceful civil rights protests supporting the Black Lives Matter movement.[8] In Baltimore, face recognition technology was used in real time to target people who were protesting after the death of Freddie Gray, with law enforcement scanning the crowd to identify individuals with outstanding warrants for unrelated offenses, and arresting them on site.[9] When

---

[3] Sam Biddle and Murtaza Hussain, "Hacked Documents: How Iran Can Track And Control Protesters' Phones", *The Intercept*, Oct. 28, 2022, https://theintercept.com/2022/10/28/iran-protests-phone-surveillance/.
[4] Khari Johnson, "Iran to use facial recognition to identify women without hijabs", *Ars Technica*, Jan. 11, 2023, https://arstechnica.com/tech-policy/2023/01/iran-to-use-facial-recognition-to-identify-women-without-hijabs/.
[5] Alfred Ng, "How China uses facial recognition to control human behavior", *CNET*, Aug. 11, 2020, https://www.cnet.com/news/politics/in-china-facial-recognition-public-shaming-and-control-go-hand-in-hand/ ("The punishing of these minor offenses is by design, surveillance experts said. The threat of public humiliation through facial recognition helps Chinese officials direct over a billion people toward what it considers acceptable behavior, from what you wear to how you cross the street").
[6] Paul Mozur, "One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority", *The New York Times*, Apr. 14, 2019, https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html.
[7] Paul Mozur, "In Hong Kong Protests, Faces Become Weapons", *The New York Times*, July 26, 2019, https://www.nytimes.com/2019/07/26/technology/hong-kong-protests-facial-recognition-surveillance.html; Lena Masri, "Facial recognition is helping Putin curb dissent with the aid of U.S. tech", *Reuters*, Mar. 28, 2023, https://www.reuters.com/investigates/special-report/ukraine-crisis-russia-detentions/; Stephen Kafeero, "Uganda is using Huawei's facial recognition tech to crack down on dissent after anti-government protests", *Quartz*, Nov. 27, 2020, https://qz.com/africa/1938976/uganda-uses-chinas-huawei-facial-recognition-to-snare-protesters.
[8] Joanne Cavanaugh Simpson and Marc Freeman, "South Florida police quietly ran facial recognition scans to identify peaceful protestors. Is that legal?", *Sun Sentinel*, June 26, 2021, https://www.sun-sentinel.com/2021/06/26/south-florida-police-quietly-ran-facial-recognition-scans-to-identify-peaceful-protestors-is-that-legal/.
[9] Kevin Rector and Alison Knezevich, "Social media companies rescind access to Geofeedia, which fed information to police during 2015 unrest", *The Baltimore Sun*, Oct. 11, 2016, https://www.baltimoresun.com/news/crime/bs-md-geofeedia-update-20161011-story.html.

face recognition is used in this way, it violates people's rights to freedom of expression and peaceful assembly. Congress must act to rein it in.

Facial recognition technology is becoming more widely available and cheaper to use. A study by Georgetown's Center on Privacy and Technology published in 2016 showed that at least one in four state and local law enforcement agencies had access to facial recognition – and that was seven years ago.[10] Research suggests that the FBI conducts thousands of scans per month, matched against reference databases of hundreds of millions of photos.[11] Several years ago, Americans were shocked to learn about the practices of the private company Clearview AI, which claims to have scraped over 20 *billion* photographs from the internet to power its face recognition systems.[12] Clearview has now been used by over 3000 federal, state and local law enforcement agencies in the United States to provide facial recognition services.[13]

Policymakers should treat facial recognition as a priority because it is a double-edged sword: Facial recognition is dangerous when it works poorly, and dangerous in an entirely different way when it works well. States have begun to respond to this threat, with over a dozen enacting meaningful limits and some jurisdictions banning the technology.[14] It is critical that Congress act as well. As our nation considers its approach to governing AI, this is an area where Congress could draw a clear contrast to autocratic regimes, demonstrating America's commitment to human rights.

The urgent need for regulation of facial recognition technology is clear. Facial recognition misidentifications have already caused numerous innocent people to be wrongfully arrested and jailed. Most recently, Randel Reid was held for six days in a Georgia jail because a facial recognition system misidentified him,[15] the latest in a series of known cases.[16] Because of police overreliance on AI, these individuals faced indignity, deprivation of liberty, and lasting harms such as loss of employment, steep legal fees, and mental trauma.[17] And since police use of facial recognition is often hidden,[18] these incidents likely represent just the tip of the iceberg.[19]

---

[10] The Perpetual Line-Up: Unregulated Police Face Recognition in America, Georgetown Law Center on Privacy and Technology, Oct. 18, 2016, https://www.perpetuallineup.org/.
[11] *Id.*; *see also* Charlie Osborne, "FBI, ICE plunder DMV driver database 'gold mine' for facial recognition scans", *ZDNET*, July 8, 2019, https://www.zdnet.com/article/fbi-and-ice-are-using-dmv-gold-mine-for-facial-recognition-scans/.
[12] Kashmir Hill, "Your Face is Not Your Own", *The New York Times Magazine*, Mar. 18, 2021, https://www.nytimes.com/interactive/2021/03/18/magazine/facial-recognition-clearview-ai.html.
[13] *Id.*
[14] Jake Laperruque, "Limiting Face Recognition Surveillance: Progress and Paths Forward", Center for Democracy & Technology, Aug. 23, 2022, https://cdt.org/insights/limiting-face-recognition-surveillance-progress-and-paths-forward/.
[15] Kashmir Hill and Ryan Mac, "'Thousands of Dollars for Something I Didn't Do'", *The New York Times*, Mar. 31, 2023, https://www.nytimes.com/2023/03/31/technology/facial-recognition-false-arrests.html.
[16] Khari Johnson, "How Wrongful Arrests Based on AI Derailed 3 Men's Lives", *WIRED*, Mar. 7, 2022, https://www.wired.com/story/wrongful-arrests-ai-derailed-3-mens-lives/.
[17] *Id.*; *see also* Elaisha Stokes, "Wrongful arrest exposes racial bias in facial recognition technology", *CBS News*, Nov. 19, 2020, https://www.cbsnews.com/news/detroit-facial-recognition-surveillance-camera-racial-bias-crime/; Kashmir Hill, "Another Arrest, and Jail Time, Due to a Bad Facial Recognition Match", *The New York Times*, Dec. 29, 2020, https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html.
[18] Khari Johnson, "The Hidden Role of Facial Recognition Tech in Many Arrests", *WIRED*, Mar. 7, 2022, https://www.wired.com/story/hidden-role-facial-recognition-tech-arrests/; Jennifer Valentino-DeVries, "How the Police Use Facial Recognition, and Where It Falls Short", *The New York Times*, Jan. 12, 2020, https://www.nytimes.com/2020/01/12/technology/facial-recognition-police.html.
[19] Disturbingly, some facial recognition misidentifications likely have resulted in prison time for innocent persons, either wrongfully convicted or pressured to accept a plea bargain out of fear of long sentences or extended time in pretrial detention.

Misidentification stems from a range of causes. Most facial recognition systems display algorithmic bias; studies have repeatedly shown propensity to misidentify people of color and women at higher rates than white people and men.[20] Software settings and nature of use impact accuracy as well. Many law enforcement agencies, including the FBI, set their systems to return several potential matches for *every* facial recognition scan even if the "confidence threshold"—meaning the required level of certainty to list an individual as a possible match—is unreliably low.[21] Law enforcement also regularly uses dubious methods to alter or replace images before scanning, from using CGI to artificially fill in uncaptured portions of a face, to replacing photos entirely with a composite sketch or celebrity look alike.[22] Finally, accuracy can vary significantly based on image quality: Lighting, photo resolution, distance, camera angle, and facial obstructions can all have a major impact on whether facial recognition returns accurate matches.[23] This is critical because even if algorithmic bias were solved, and responsible settings and use parameters were employed, varying image quality will always cause misidentification risk.

Just as serious as misidentifications are the dangers of accurate facial recognition being used for surveillance. The examples I shared previously from Iran, China, Russia, Uganda – and at least three U.S. cities – shows how easily face recognition technology can impinge on people's rights to express themselves through protest and to peacefully assemble. Facial recognition could be employed to monitor, catalog, and engage in disparate targeting of individuals participating in a variety of sensitive or constitutionally protected activities, such as attending a political rally, going to a house of worship, purchasing a firearm from a licensed shop, or visiting a medical clinic. Absent strong limits, law enforcement authorities could misuse AI technology to track and catalog individuals' most sensitive activities with little effort, and on an unprecedented scale. The U.S. must show leadership by curtailing such a direct assault on civil liberties.

Given the range of risks facial recognition poses to civil rights and civil liberties, there is not a silver bullet policy solution: lawmakers need to enact a broad set of safeguards to prevent harm,

---

[20] Joy Buolamwini and Timnit Gebru (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, *Fairness, Accountability and Transparency, Proceedings of Machine Learning Research* 81:77-91. http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf; Patrick Grother, Mei Ngan, and Kayee Hanaoka (Dec. 2019). *Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects*, National Institute Of Science and Technology. https://doi.org/10.6028/NIST.IR.8280.

[21] Kimberly J. Del Greco, "Facial Recognition Technology: Ensuring Transparency in Government Use", Federal Bureau of Investigation, June 4, 2019, https://www.fbi.gov/news/testimony/facial-recognition-technology-ensuring-transparency-in-government-use; Drew Harwell, "Oregon became a testing ground for Amazon's facial-recognition policing. But what if Rekognition gets it wrong?", *The Washington Post*, April 30, 2019, https://www.washingtonpost.com/technology/2019/04/30/amazons-facial-recognition-technology-is-supercharging-local-police/ ("But deputies here are not shown that search-confidence measurement when they use the tool. Instead, they are given five possible matches for every search, even if the system's certainty in a match is far lower").

[22] James O'Neill, "How Facial Recognition Makes You Safer", *The New York Times*, June 9, 2019, https://www.nytimes.com/2019/06/09/opinion/facial-recognition-police-new-york-city.html; Clare Garvie, "Garbage In, Garbage Out | Face Recognition on Flawed Data", Georgetown Law Center on Privacy & Technology, May 16, 2019, https://www.flawedfacedata.com/ ("One detective from the Facial Identification Section (FIS), responsible for conducting face recognition searches for the NYPD, noted that the suspect looked like the actor Woody Harrelson …. A Google image search for the actor predictably returned high-quality images, which detectives then submitted to the face recognition algorithm in place of the suspect's photo.")

[23] The Constitution Project's Task Force on Facial Recognition Surveillance and Jake Laperruque, "Facing the Future of Surveillance", Project on Government Oversight, Mar. 4, 2019, https://www.pogo.org/report/2019/03/facing-the-future-of-surveillance.

both from misidentifications and misuse.[24] The Center for Democracy & Technology views the following measures as key to effectively regulating law enforcement use of facial recognition:

1) **A warrant rule:** Law enforcement use of facial recognition should require obtaining a warrant from a judge, based on probable cause that the individual to be scanned committed a crime.[25] Warrants are a fundamental privacy safeguard and key to preventing abuse, notably using facial recognition to identify, catalog, and target individuals engaged in lawful and sensitive activities, such as protests.
2) **A serious crime limit:** Face recognition technology should be restricted for use only in investigating serious offenses.[26] Such limitations would prevent selective targeting and prosecution, as well as prevent misidentifications in scenarios least likely to receive due scrutiny: the investigation and prosecution of low-level crimes.
3) **Notification for arrested individuals:** Law enforcement should not be allowed to routinely hide their use of facial recognition from defendants and the broader public.[27] This common practice undermines defendants' due process rights, and prevents examination of errors and other meaningful oversight.
4) **Prohibiting overreliance on matches:** Police should not be permitted to use facial recognition as the sole basis for arrests or other police actions. Given that the technology's accuracy varies significantly based on a range of factors, independent investigative work is essential.
5) **Prohibiting untargeted scans:** Facial recognition technology may soon focus on untargeted scans—whereby every individual passing through a video feed is identified with facial recognition—but this method is far too unreliable for law enforcement use. Pilot programs have produced false positives of 81 to 96 percent.[28] Even if these extreme error rates were to improve, such a use of face recognition technology would constitute unacceptable dragnet surveillance that should not be deployed.
6) **Testing and accuracy standards:** Any law enforcement use of facial recognition should require that software be subject to independent testing and meet accuracy standards. Testing should focus on live field conditions that replicate investigative use, and accuracy standards should limit use to algorithms with highest overall accuracy and that display no variance based on demographic traits.

---

[24] While our recommendations focus on safeguards and limits for law enforcement use of facial recognition, it is important to acknowledge that many privacy, civil rights, and civil liberties groups—including CDT—have called for a moratorium on facial recognition, or for its use by law enforcement to be banned entirely. Some local face recognition laws have taken this approach. CDT supports enacting a moratorium while evaluating proper restrictions and safeguards as providing the strongest protections for civil rights and civil liberties. *See, e.g.*, LDF Letter re: July 13, 2021 Subcommittee on Crime, Terrorism, and Homeland Security Hearing on Law Enforcement Use of Facial Recognition Technology, https://www.naacpldf.org/wp-content/uploads/2021.07.20-LDF-Statement-on-Law-Enforcement-U_Emily-Fisher-1.pdf.

[25] This should include sensible limited exceptions, such as identifying victims and incapacitated persons.

[26] A serious crime limit has been used for over 50 years to prevent wiretap surveillance from becoming pervasive. *See* 18 U.S.C. § 2516.

[27] Khari Johnson, "The Hidden Role of Facial Recognition Tech in Many Arrests", *WIRED*, Mar. 7, 2022, https://www.wired.com/story/hidden-role-facial-recognition-tech-arrests/; Jennifer Valentino-DeVries, "How the Police Use Facial Recognition, and Where It Falls Short", *The New York Times*, Jan. 12, 2020, https://www.nytimes.com/2020/01/12/technology/facial-recognition-police.html.

[28] Lizzie Dearden, "Facial recognition wrongly identifies public as potential criminals 96% of time, figures reveal", *The Independent*, May 7, 2019, https://www.independent.co.uk/news/uk/home-news/facial-recognition-london-inaccurate-met-police-trials-a8898946.html; Rachel England, "UK police's facial recognition system has an 81 percent error rate", *Engadget*, July 4, 2019, https://www.engadget.com/2019-07-04-uk-met-facial-recognition-failure-rate.html.

The adoption of face recognition laws by over a dozen states[29] demonstrates an emerging consensus for regulating this surveillance. Unfortunately, thus far Congress has placed no limits on facial recognition, leaving this powerful technology unrestricted. Last year a bill was introduced in the House, H.R. 9061, The Facial Recognition Act, that included many of the recommendations listed above, and that the Center for Democracy & Technology endorsed.[30] We encourage Congress to act with urgency to place safeguards on this form of AI surveillance, and focus on the policies described above.

Generative AI, Elections & Democratic Discourse

Turning to my second area of focus, rapid advances in generative AI are spurring creativity and innovation, but also raise significant threats for human rights. Already there have been instances showing the professional, reputational and potential physical harms that may arise when people rely on generated results as accurate, not accounting for the likelihood of "hallucinations", or mistaken results.[31] Generative AI tools are likely to exacerbate fraud, as tools make it easier to quickly generate massive amounts of convincing text, as well as personalized scams, or to trick people by impersonating a familiar voice.[32] Deepfakes – videos or images that have been digitally manipulated to misrepresent the voice and likeness of another person – can misrepresent public figures or events in a way that threatens elections, national security, and general public order.[33] Deepfakes can also be used to defraud, harass, and extort people.[34] None of these harms is new, but they are made cheaper, faster, and more effective by the ease, speed and widespread accessibility of generative AI tools.

The threats to elections and democratic discourse are particularly worth highlighting. In previous elections, operatives used robocalls to spread incorrect information about mail-in voting in an effort to suppress Black voter turnout,[35] and deceptive text messages to spread intentionally misleading voting instructions for a Kansas ballot initiative in 2022.[36] It is easy to imagine bad actors using AI to exponentially grow and personalize voter suppression or other targeting efforts, increasing their harmful impact. Today, consumers can often spot a scam email, text or robocall because it uses non-personalized language and there may be grammatical

[29] Jake Laperruque, "Limiting Face Recognition Surveillance: Progress and Paths Forward", Center for Democracy & Technology, Aug. 23, 2022, https://cdt.org/insights/limiting-face-recognition-surveillance-progress-and-paths-forward/.
[30] Jake Laperruque, "The Facial Recognition Act: A Promising Path to Put Guardrails on a Dangerously Unregulated Surveillance Technology", *Lawfare*, Nov. 1, 2022, https://www.lawfareblog.com/facial-recognition-act-promising-path-put-guardrails-dangerously-unregulated-surveillance-technology.
[31] Karen Weise and Cade Metz, "When A.I. Chatbots Hallucinate", *The New York Times*, May 1, 2023, https://www.nytimes.com/2023/05/01/business/ai-chatbots-hallucination.html.
[32] Steve Mollman, "Scammers are using voice-cloning A.I. tools to sound like victims' relatives in desperate need of financial help. It's working", *Fortune*, Mar. 5, 2023, https://fortune.com/2023/03/05/scammers-ai-voice-cloning-tricking-victims-sound-like-relatives-needing-money/.
[33] Shannon Bond, "Fake viral images of an explosion at the Pentagon were probably created by AI", *NPR*, May 22, 2023, https://www.npr.org/2023/05/22/1177590231/fake-viral-images-of-an-explosion-at-the-pentagon-were-probably-created-by-ai; David Klepper and Ali Swenson, "AI presents political peril for 2024 with threat to mislead voters", *AP News*, May 14, 2023, https://apnews.com/article/artificial-intelligence-misinformation-deepfakes-2024-election-trump-59fb51002661ac5290089060b3ae39a0.
[34] *See e.g.,* Henry Ajder, Giorgio Patrini and Francesco Cavalli, "Automating Image Abuse: Deepfake bots on Telegram", *Sensity*, Oct. 2020 (deepfake bots on Telegram digitally "undress" more than 100,000 women on the platform); Thomas Brewster, "Fraudsters Cloned Company Director's Voice In $35 Million Heist, Police Find", *Forbes*, Oct. 14, 2021, https://www.forbes.com/sites/thomasbrewster/2021/10/14/huge-bank-fraud-uses-deep-fake-voice-tech-to-steal-millions/?sh=7d29a3f87559 (audio deepfake of executives' voices used to steal millions of dollars from companies).
[35] Christine Chung, "They Used Robocalls to Suppress Black Votes. Now They Have to Register Voters.", *The New York Times*, Dec. 1, 2022, https://www.nytimes.com/2022/12/01/us/politics/wohl-burkman-voter-suppression-ohio.html.
[36] Isaac Stanley-Becker, "Misleading Kansas abortion texts linked to Republican-aligned firm", *The Washington Post*, Aug. 2, 2022, https://www.washingtonpost.com/politics/2022/08/02/kansas-abortion-texts/.

or language errors (or, in the case of robocalls, a notably automated voice). Generative AI tools will make it easier to create tailored, accurate, realistic messages that draw victims in.

Generated images can also twist public understanding of political figures and events. Recordings of public figures' voices have been manipulated to trick senior government officials into thinking they are speaking with government leaders.[37] Videos and images have been digitally altered to make public officials appear incompetent, compromised, or to misrepresent their policy positions.[38] Experts have warned how deepfakes, which are difficult to authenticate or rebut, could impact an election in the closing days of voting, when there is little time to set the record straight, or before a debate.[39] More generally, the growth of inauthentic content makes it harder for people to know what news and content they can trust, such that even authentic content is undermined. Journalists, whistleblowers, and human rights defenders are experiencing these effects already, facing higher hurdles than ever before to establish and defend their credibility.[40]

While the rise of affordable generated content poses new threats to public discourse, policy interventions must be approached with care. This is because there are many legitimate reasons why people use software to generate and alter content: from laypeople and artists using AI to make creative works; to people engaging in parody; actors being de-aged in a movie; voices being sampled for a music track; or researchers altering images of North American and European cities to show what they would look like if they faced the same bombardment as the cities attacked in the Syrian war.[41] Barring or heavily restricting such activities would harm free expression, creativity and innovation, and quickly run afoul of the First Amendment.

Efforts to restrict or condition the distribution of generative images may also suppress protected expressive activities. To give one example, in recent years a number of companies and stakeholders have come together in the Content Authenticity Initiative, an impressive undertaking that allows photographers and other content creators to attach immutable provenance signals showing the authenticity of their work (such as details of the image's creator, date/time/location, tracked edits and more).[42] This is a creative solution to help newspapers, human rights watchdogs and others reassure the public about the authenticity and provenance of images they create and display. But *mandating* the use of such an authenticity standard (or

---

[37] *See e.g.,* Bobby Allyn, "Deepfake video of Zelenskyy could be 'tip of the iceberg' in info war, experts warn", *NPR*, Mar. 16, 2022, https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia (the minute long deepfake video "shows a rendering of the Ukrainian president appearing to tell his soldiers to lay down their arms and surrender the fight against Russia"); Philip Oltermann, "European politicians duped into deepfake video calls with mayor of Kyiv", *The Guardian*, Jun. 25, 2022, https://www.theguardian.com/world/2022/jun/25/european-leaders-deepfake-video-calls-mayor-of-kyiv-vitali-klitschko.

[38] *See e.g.,* Hannah Denham, "Another fake video of Pelosi goes viral on Facebook", *The Washington Post,* Aug. 3, 2020, https://www.washingtonpost.com/technology/2020/08/03/nancy-pelosi-fake-video-facebook/ (video depicts Pelosi slurring her speech and appearing intoxicated"); Alexandra Ulmer and Anna Tong, "Deepfaking it: America's 2024 election collides with AI boom", *Reuters*, May 30, 2023, https://www.reuters.com/world/us/deepfaking-it-americas-2024-election-collides-with-ai-boom-2023-05-30/; Zeke Miller, "Rubio Campaign Fires Back at Cruz Over Photoshopped Image", *Time*, Feb. 18, 2016, https://time.com/4229092/marco-rubio-ted-cruz-photoshop/. While running for re-election in 2019, Houston's mayor said a critical ad ran by a fellow candidate broke a Texas law that bans certain misleading political deepfakes. Ivory Hecker, "Mayor Turner calls for criminal investigation of Tony Buzbee's attack ad", *Fox 26 Houston*, Oct. 17, 2019, https://www.fox26houston.com/news/mayor-turner-calls-for-criminal-investigation-of-tony-buzbees-attack-ad.

[39] James Bickerton, "Deepfakes Could Destroy the 2024 Election", *Newsweek*, Mar. 24, 2023, https://www.newsweek.com/deepfakes-could-destroy-2024-election-1790037.

[40] Sam Gregory, "Tracing trust: Why we must build authenticity infrastructure that works for all", *Witness*, May 2020, https://blog.witness.org/2020/05/authenticity-infrastructure/.

[41] Tiffany Hsu, "As Deepfakes Flourish, Countries Struggle With Response", *The New York Times*, Jan. 22, 2023, https://www.nytimes.com/2023/01/22/business/media/deepfake-regulation-difficulty.html.

[42] *See* Content Authenticity Initiative, https://contentauthenticity.org/.

prohibiting the distribution of materials without such standards) would be deeply problematic, because it would suppress the posting and sharing of lawful images whose creators lacked the resources or awareness to use a provenance tool, who face safety risks if their work can be traced back to them, or who simply do not want to do so.

The challenges of regulating deepfakes does not mean policymakers must sit idle. To the contrary, there are concrete steps Congress can take to increase transparency and accountability in the design, development and use of generative AI tools, as well as appropriations provisions, oversight of relevant federal agencies, and steps such as hearings, convenings, and/or the creation of a Commission to highlight best practices and novel innovations to address potential harms.

1) **Mandating transparency & disclosures of AI risks.** Several legislative proposals introduced last Congress seek to increase the accountable design and transparency of AI systems, including the Algorithmic Accountability Act, and the algorithmic impact assessment provision of the bipartisan American Data Privacy & Protection Act. These measures were drafted before the wide-scale public release of generative AI systems, but their principles lay an important foundation for future work.

   As a starting point, Congress could require the developers of AI systems that can be used in high-risk settings to disclose how their tools are developed and designed, to test them using frameworks based on principles such as those set out in the Blueprint for an AI Bill of Rights and the NIST AI Risk Management Framework, and to share the analysis of those tests with an outside regulator (with some version made available for the public and for independent researchers, balancing concerns about the potential privacy and safety aspects of such disclosures). Such steps would increase transparency and support meaningful public dialogue about how tools are developed and governed. They would also normalize the principle that companies designing and deploying AI tools *must* analyze and document how they work, identify potential risks, and disclose the steps they have taken to mitigate those risks. Such legislation would establish an essential baseline, and need not foreclose potential legislation on minimum design and safety standards, the specific regulation of highly capable foundation models, or further steps to address other high-risk AI uses.

2) **Examining how existing criminal and civil laws map onto harms created by new tools, and filling gaps.** In some instances, the appropriate framework to address harms created by generative AI (and other AI systems) may be litigation under existing laws. For example, people who use AI to perpetrate scams could be prosecuted for fraud, extortion, or harassment; face investigation by the Federal Trade Commission for unfair and deceptive trade practices or the Federal Elections Commission for violating campaign laws; or face civil litigation for claims such as fraud, intentional infliction of emotional distress, harassment, defamation and intellectual property violations. Congress should monitor whether these existing legal frameworks adequately address emerging harms.[43]

[43] Four federal agencies recently announced their efforts to enforce existing laws to protect the American public from AI-related harms. Other agencies should take similar steps, and Congressional committees of relevant jurisdiction can support these efforts to understand how existing

In assessing liability, courts will have to tackle the complex question of whether and when developers of generative AI tools bear legal liability for the content those tools produce. Courts will have to consider whether the content generated by an AI tool is properly considered to be the speech of the user who prompted its creation, or something partially or wholly created or developed by the AI tool itself. This will likely differ depending on the fact pattern: for example, whether a user inputted specific prompts aiming to generate the content that gave rise to litigation, such as soliciting a list of crimes committed by a private individual and publishing that list with reckless disregard for whether the information was true, or whether the AI tool was the source of the content giving rise to litigation, , such as making up dangerously incorrect medical advice in response to a query. In addition to statutes and case law regarding intermediary liability protections, courts will need to consider a range of common law principles from across civil and criminal law, including standards for aiding and abetting liability, and questions of knowledge and intent for both the user and the developer (and, if different, the deployer) of the tool. Companies will need to point to content policies and technical safeguards they have in place to mitigate foreseeable misuses and other harms.

As courts grapple with these and other complex issues, Congress can shine a light and drive public discourse — and then act as appropriate to fill in the gaps. This could include hearings and reports by Congressional committees in their areas of jurisdiction, commissioning reports by the GAO or federal agencies, or, more formally, the creation of an expert Commission to advance such work.[44]

3) **Advancing best practices for responsible design and governance of generative AI systems**. There is an urgent need for companies developing generative AI systems to develop robust safety processes and other governance measures, as many of their CEOs have themselves publicly declared.[45] This can include steps ranging from well-developed content policies and technical safeguards that limit the creation of certain high-risk content or uses of the technology;[46] robust pre- and post-release testing to identify and address bias and potential harms; improved interfaces, labeling and product descriptions to better educate

---

laws map onto novel fact patterns. *See* Joint Statement on Enforcement Efforts Against Discrimination and Bias in Automated Systems, Apr. 25, 2023, https://www.ftc.gov/system/files/ftc_gov/pdf/EEOC-CRT-FTC-CFPB-AI-Joint-Statement%28final%29.pdf.

[44] *See, e.g.,* Deepfake Task Force Act, S.2559, 117th Cong. (2021-2022); American Data Privacy & Protection Act of 2022 (ADPPA), H.R. 8152, 117th Cong. (2021-2022). These proposals both focus on creating a task force (or in the case of the ADPPA, mandating annual reporting by the Commerce Department) on the uses and harms of deepfakes and advancements in deepfake detection technology. But a Commission could also be charged with reporting on and assessing existing legal frameworks for addressing and seeking redress for other harms.

[45] *See e.g.,* Sam Altman, Oversight of A.I.: Rules for Artificial Intelligence Hearing before the U.S. Senate Committee on the Judiciary Subcommittee on Privacy, Technology, & the Law, 118th Cong. (2023), https://www.judiciary.senate.gov/committee-activity/hearings/oversight-of-ai-rules-for-artificial-intelligence; Sundar Pichai, "Why Google thinks we need to regulate AI", *Financial Times*, Jan. 20, 2020, https://www.ft.com/content/3467659a-386d-11ea-ac3c-f68c10993b04 (CEO of Google stating that "there is no question in [his] mind that artificial intelligence needs to be regulated"); Brad Smith, "Meeting the AI moment: advancing the future through responsible AI", Microsoft, Feb. 2, 2023, https://blogs.microsoft.com/on-the-issues/2023/02/02/responsible-ai-chatgpt-artificial-intelligence/ (Vice Chair & President of Microsoft calling for effective AI regulations that "center on the highest risk applications and be outcomes-focused and durable").

[46] For example, OpenAI claims that its image generator DALL-E cannot create images of public figures, and that it restricts any "scaled" usage of its products for political purposes, such as the use of its AI to send out mass personalized emails to constituents. Reporters testing these claims have found significant exceptions and workarounds. Robust, well-tested and publicly disclosed content policies form an important aspect of safety testing. Alexandra Ulmer and Anna Tong, "Deepfaking it: America's 2024 election collides with AI boom", *Reuters*, May 30, 2023, https://www.reuters.com/world/us/deepfaking-it-americas-2024-election-collides-with-ai-boom-2023-05-30/.

users about the systems' limitations and risks of inaccurate results;[47] safeguarding systems against security threats, and more.

Governments in different countries are pressing companies on what these steps should look like.[48] Whether or not these steps are ripe for legislation, Congress can play a role in driving forward these efforts – and, most critically, ensuring they are not taking place behind closed doors with only companies in attendance, but instead with meaningful participation from civil society and independent sources of expertise.

4) **Scaling agencies' capacity to address deepfakes and boost authentic sources of information.** It has long been said that the best remedy to combat undesirable speech is counterspeech[49] – but in our cacophonous information ecosystem, it takes work for counterspeech to be effective. There are steps policymakers can take to mature the systems that can help individuals better understand content authenticity and identify reliable sources of information. As one step, the government could increase funding and other efforts to support the development of technologies that assist in deepfake detection.[50] Policymakers could also support and foster awareness of voluntary efforts to authenticate content, funding research projects through the National Science Foundation and other programs, or raising awareness of private sector efforts to encourage the quick development of such work.[51]

Critically, Congress and the Administration should significantly ramp up efforts to equip key institutions so they can identify and debunk manipulated content that threatens national security, financial markets, election administration, public health and similar priority areas. The bipartisan Deepfake Task Force Act proposed last Congress provides a good bipartisan foundation from which to start. That measure directed the creation of a task force comprised of government and non-government experts to "investigate the feasibility of, and obstacles to, developing and deploying standards and technologies for determining digital content provenance", and created "a formal mechanism for interagency coordination and information sharing to facilitate the creation and implementation of a national strategy to address the growing threats posed by digital content forgeries."[52]

---

[47] Michal Luria, "Your ChatGPT Relationship Status Shouldn't Be Complicated", *WIRED*, Apr. 11, 2023, https://www.wired.com/story/chatgpt-social-roles-psychology/.

[48] Ryan Browne, "With ChatGPT hype swirling, UK government urges regulators to come up with rules for A.I.", *CNBC*, Mar. 29, 2023, https://www.cnbc.com/2023/03/29/with-chatgpt-hype-swirling-uk-government-urges-regulators-to-come-up-with-rules-for-ai.html; Ryan Browne, "Europe takes aim at ChatGPT with what might soon be the West's first A.I. law. Here's what it means", *CNBC*, May 15, 2023, https://www.cnbc.com/2023/05/15/eu-ai-act-europe-takes-aim-at-chatgpt-with-landmark-regulation.html. In the U.S., the White House issued the AI Bill of Rights in October 2022 and the National Institute of Standards and Technology (NIST) followed in January 2023 with an AI Risk Management Framework, and officials have spoken about ways in which these map onto the risks posed by generative AI. *See* Blueprint for an AI Bill of Rights, https://www.whitehouse.gov/ostp/ai-bill-of-rights/; National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

[49] *Whitney v. Cal.*, 274 U.S. 357, 377 (1927) (Brandeis, J., concurring) ("If there be time to expose through discussion the falsehood and fallacies, to avert the evil by the processes of education, the remedy to be applied is more speech, not enforced silence.").

[50] *See, e.g.*, IOGAN Act, Pub. L. No. 116-258 (2020), directing the National Science Foundation and the National Institute of Standards and Technology (NIST) to support research on generative adversarial networks. The proposed American Data Privacy & Protection Act of 2022 would have required the Secretary of Commerce to publish an annual report on common sources of digital content forgeries, an assessment of the uses, applications and harms of digital content forgeries, and an analysis of the methods and standards available for detection and counter-measures such as labeling. American Data Privacy & Protection Act of 2022, H.R. 8152, Section 305, 117th Cong. (2021-2022).

[51] Shirin Ghaffary, "What will stop AI from flooding the internet with fake images?, *Vox*, Jun. 3, 2023, https://www.vox.com/technology/23746060/ai-generative-fake-images-photoshop-google-microsoft-adobe.

[52] Section 5709 of the National Defense Authorization Act of 2020 also took steps to improve government agency awareness and competency to address deepfakes. It directed the Director of National Intelligence to produce a report on the technological capabilities of foreign actors with

Capacity-building efforts could also include funding training, resources and using oversight pressure to ensure public institutions take steps to best earn public trust when they speak out. To give one simple example, research by my organization, the Center for Democracy & Technology, revealed that only in 1 in 4 official election websites uses the trusted ".gov" domain managed by DHS, while other election officials use ".com" domains that can be easily spoofed. The result is to undermine the role of such websites as a source for people to access trusted information about the administration of elections. Funding, education and oversight could help election officials address this simple vulnerability.

Conclusion

The examples of face recognition and misleading information about elections show two very different ways in which AI is already impacting Americans' human rights and the structure of our democracy. Critically, these examples show that there are concrete steps policymakers can take, today, to address the potential harms that can arise from certain uses of AI. As commentators around the world assess the existential threats posed by AI systems, it is important to remember that existential threats can also include threats to the fabric of society: undermining individual rights, equality and economic mobility, and an informed public discourse that is the bedrock of a functioning democracy. On many of these issues, there are steps that technology companies, regulatory agencies and Congress can take right now to address and reduce AI-driven harms. Thank you for the opportunity to share these thoughts today.

---

respect to "machine-manipulated media, machine generated text, generative adversarial networks, and related machine-learning technologies", and analysis of the counter-technologies that have been or could be developed and deployed to address such uses, among other factors. National Defense Authorization Act of 2020, Pub. L. No. 116-92 (2019).

Subcommittee on Human Rights and the Law

Hearing: "Artificial Intelligence and Human Rights"

Written Statement of Aleksander Mądry[1]

June 13th, 2023

Chairman Ossoff, Ranking Member Blackburn and Members of the Committee, thank you for inviting me to testify. Much has already been said and written about how AI may transform society, both about the opportunities and risks—from AI's potential to enhance our productivity, creativity, and overall quality of life to its ability to perpetuate discrimination, drive economic inequality, and pose an existential risk.

I will not reprise those conversations here. Instead, I will focus my testimony on one issue that I find particularly salient, time-sensitive and extremely worrisome: *how AI could erode central tenets that enable our society to function, including our ability to carry out democratic decision-making.*

Specifically, I will discuss how AI is poised to fundamentally transform mechanisms for the dissemination and understanding of information, and the unsettling implications of those changes. I will also sketch out what could be done to mitigate these emerging risks.

How will AI transform the information ecosystem?

Changes in information technologies—whether the invention of the printing press, the advent of e-mail, or the emergence of social media—do not just make information more accessible, they fundamentally change the dynamics of information sharing and acquisition. While we are still dealing with the transformations in this space brought to us by email and social media, there is already a new transformation afoot—a transformation fueled by recent developments in AI that is likely to be more consequential than anything we have been experiencing recently.

With the advent of AI—especially the newest wave of generative AI—anyone who can use a chatbot is in a position to become a "trusted source"—a *highly personalized* source, in fact. Indeed, as more of what we see becomes generated and disseminated by AI, the lines between humans and bots are becoming blurred. We need to start to be more wary than ever about how information reaches us, its trustworthiness and its ability to persuade us.

More precisely, AI is changing the information-delivery landscape in three key ways:

(a) It enables the creation of content—written text, photos and, soon, videos—that seems extremely realistic.

(b) The language produced by Large Language Models (LLMs) like chatGPT or Google Bard can seem natural and highly persuasive, in no small part since we are wired to believe that such speech can come only from humans.

(c) It makes the creation of such content cheap and broadly accessible—even to parties with little if any technical expertise—making it frighteningly easy to deploy it at scale.

[1] I have recently started a professional leave from MIT, which I am spending at OpenAI. I am providing this testimony solely in my personal capacity and as an MIT faculty. I am not in any way representing OpenAI.

We are already seeing early adoption of generative AI in our information sphere, from art [10], to copywriting [7], to political ads [13], but these are just a tip of the iceberg. We will see much, much more very soon. The onset of this technology brings with it a whole spectrum of risks and potential harms. I will highlight just a few of them below.

Enhancing "traditional" cybercrime. One immediate impact of the newest wave of generative AI is that "traditional" spam and phishing campaigns are even easier to conduct. What previously required careful photo editing and writing (as well as some non-trivial human involvement) now only requires a few clicks. The recent use of an AI-generated fake image of a fire near the Pentagon is just one illustration of that [9].

Also, the fact that generative AI can convincingly impersonate a human online poses a fundamental challenge to our existing mechanisms for protecting our information infrastructure, public discourse and governance. After all, the bot detection and moderation algorithms that our online discussion platforms use—whether they be Internet forums, newspaper comment sections, or Twitter—tend to rely on some kind of "prove that you are human" tests. How will these platforms cope with malicious parties that can field swarms of sophisticated, AI-driven bots that are able to breeze through such tests?

"Spear-phishing" and personalized blackmail. The enhancement of the "traditional" deception is, however, just the beginning. AI's unique ability to create content that is both convincing and personalized means that, for example, phishing will no longer need to involve generic emails sent out to thousands of recipients, hoping someone will get duped. Instead, we will have "spear-phishing," where both the message *and* the whole conversation that ensues are *fully automated* and *customized* to you.

In fact, there is a very real possibility that a new kind of blackmail scheme will emerge. In such a scheme, someone's photo from social media is edited to depict them in a compromising situation, and then they are threatened that the edited photo will be made public unless they pay up. How many of us would not pay to simply make the problem go away? Thanks to AI these kinds of schemes can now be executed (again) *fully automatically*, *cheaply* and *at scale*.

In addition—as one of the other witnesses has experienced herself [3]—the AI-fueled ability to impersonate the voice of just about any person enables a whole new array of scams [12]. As the ability to generate video with AI improves, other troubling possibilities such as targeted AI-generated explicit content [2] will become an even more acute problem too.

Personalized persuasion at scale. This expansion of the cybercrime toolkit is hardly the only worrisome consequence though. Indeed, AI is bound to transform how we think about any information campaign—be it ideological, political or commercial. Specifically, such campaigns will no longer need to rely solely on the promoted message to go viral. Instead, they can be fielded with generative AI and the promoted messaging might reach its intended audience *individually* and in a *highly personalized* manner. So, it will not be about some post that came across your social media timeline. Rather it will be about a Facebook "friend" that you met online. Friend who is actually an AI-driven agent impersonating a human. Friend that only subtly weaves in political commentary or product endorsements or any other messaging in between your engaging conversations about sports, movies or favorite food.

Similarly, instead of trying to corral a critical mass of people to campaign for a cause—whether on social media, via direct calling, or letter-writing—a single actor can field a campaign by themselves, using generative AI-driven bots in place of people. A campaign that is *equally effective* (thanks to the sophistication of these bots) but needs neither any buy-in from the broader population nor even comparable resources. As far as I know, as of now, this would all be legal too.

Automated creation of addictive content. AI doesn't just produce content that mimics reality and appears human-like and personalized—it can also make this content *personable*. There is a lot of information about our habits, preferences, hobbies and values that can be gleaned from sources such as our social media accounts. This could make interacting with AI not only attractive and persuasive but also addictive to us. After all, loneliness and an unmet need for some kind of intimacy with others are a growing problem in our society [8], and the kind of focus, "fit" and "care" such AI-driven "friends" would seem to exhibit could be extremely alluring.

This aspect of AI could (and, I hope, will) play a positive role too [1]. But imagine the power someone who is able to deploy such AI-powered agents could have over us, especially at scale. What if that power gets abused? What if these capabilities are harnessed to supercharge the "attention economy" that already drives much of our social media and online commerce? What would this mean for our productivity and long-term happiness? How do we feel about having our children being exposed to all of that?

Eroding trust in information and written (or audio-visual) records. Thanks to AI we are entering the era when *any* record could plausibly be faked. How does this affect our collective discourse as well as the legal and governance system? After all, we are a society whose foundations rely on the veracity and binding of such records—think contracts, deposition recording, or video evidence in criminal cases—and this reliance will only increase as more of our critical interactions occur in the digital sphere. How does our society adapt to such a tectonic shift?

What can we do?

The concerns I have outlined above may paint a rather bleak and, potentially, daunting landscape. But there is much we can (and should) do here. Specifically, we need a combination of technical solutions and policy actions that will reinforce each other. After all, policy can help drive the development and implementation of technical remedies, and technical innovations can, in turn, unlock new policy options. Let me describe some of these below.

Technical solutions

On the technical front, we need tools that can help humans judge the authenticity of content—to understand the extent to which it was generated by a human and/or AI. These tools can take a variety of forms (and for many of them we already have proof-of-concept prototypes):

Watermarking and deepfake detection tools. One promising idea for ensuring the authenticity of content is "watermarking"—that is, placing an imperceptible "signature" in generated content that makes clear AI was used. This watermark can then be detected by any content consumer. Researchers have developed prototypes of watermarking systems, both in the context of large language

models [4] and image generations models [14]. Much more work is needed, however, to make them sufficiently robust and then policies might be needed to drive their adoption too. Also, like all such technologies, there would likely be an "arms race"—tools will be developed to evade the watermark system and improved techniques will be needed to respond to that.

Watermarks need to be placed in documents directly by the AI providers, but there is also a line of work on detecting AI-generated content in the absence of cooperation from the developers of a given AI model [6]. Of course, this lack of cooperation makes it easier for malicious actors to thwart these detection techniques, causing the corresponding "arms race" to be much more challenging.

Protection against unauthorized AI-powered content editing. Another problem that technology can help address is unauthorized AI-powered content editing—that is, the ability to use AI-powered editing tools to manipulate content against the wishes of its creators or people depicted in it. (Think, for example, of the personalized blackmail scheme described earlier, which involved a malicious party manipulating photos the victim had published on social media.) Could we develop a way for users to protect the photos they put online, to make it impossible—or, at least, much harder—to modify using AI? It turns out that such an "immunization" capability is a possibility [11] but, again, much more work is needed.

Provenance certification techniques. Beyond detecting AI-generated content, tools may be needed to *prove* the authenticity of content. This could involve, for example, leveraging cryptographic tools to provide automatic certification of the authenticity or provenance of a given document by tracing it to the exact primary source that created it (e.g., the person who took a given photo). When such a technology is broadly available, content might be presumed to be fake unless verification proves it to be real.

However, just to reiterate: no matter how work on such tools proceeds, these tools will *not* be a panacea. They will be neither perfect nor foolproof, either—that is not technically possible. Nonetheless, these tools can provide the necessary "friction" that makes undesirable use of AI that much harder to execute and they will also create "footholds" for the policy action.

Policy solutions

As I noted above, technological approaches will need to work hand-in-hand with policy. Here are some possible policy approaches to pursue.

AI-generated content disclosure requirement. One relatively straightforward step would be to require that any consumer-facing AI-generated content be labeled as such. This kind of mandatory disclosure would, for example, likely hamper an AI-powered persuasion campaign we described above—at least, as long as this rule was abided by.

Of course, deciding the exact level of AI involvement that would trigger such a mandate—as well as the form it would need to take—would require careful deliberation. And the rules would have to be updated as the technology and the use of it evolved. In particular, it would be important to avoid the "user desensitization" effect, in which the users stop paying attention to the corresponding disclosures due to being bombarded with them at every occasion (and for trivial reasons). (Such desensitization seems to have occurred, for example, in the context of the web cookie usage disclosure

and consent requirements imposed in the European General Data Protection Regulation (GDPR) [5].)

Accelerating the use of content authenticity tools. As discussed earlier, content authenticity tools such as watermarking, deepfake detection, protection against unauthorized AI-powered editing, or provenance certification can be very useful but their effectiveness is hardly guaranteed. Even leaving aside technical questions, the efficacy of these solutions will critically depend on how broadly adopted they are. We need here a broad cooperation of the industry players that develop the relevant AI systems, so as to establish consistent expectations and standards. Policy can accelerate this process and broaden the use of such techniques, through incentives and/or mandates. After all, we don't know if market incentives will ever be sufficiently strong to drive the development and deployment of these technologies; they certainly are not enough at this point.

Client identification and suspicious activity reporting mandates. One possible approach to deterring rogue actors could be adapted from anti-money laundering laws. It would require providers of sufficiently capable AI services to implement adequate client identification mechanisms. These AI providers would then be expected to monitor the usage of the tools they supply to flag (and, potentially, block) suspicious activity as well as to report it to appropriate governmental agencies (such as FBI) or other organizations.

Advance "AI literacy" efforts. Of course, no technical solution or set of regulations will ever suffice to fully mitigate the risks AI now poses. It is thus crucial that, in addition to "email literacy" and "social media literacy," we think about promotion of "AI literacy." The public needs to understand how to judiciously interact with AI systems—and how to be on the lookout for when they are interacting with AI in the first place. This includes helping the public avoid the natural tendency to anthropomorphize AI systems. After all, AI does not reason; it merely mimics reasoning—at least as of now. We also must go from assuming that content is authentic until proven otherwise to assuming that content is fake until proven otherwise—or at the very least discounting the value of unverified content.

Overall, there is a need for a shift in the public mindset to accommodate how AI is changing the world. We thus need a decisive policy thinking on how to advance such AI literacy more intentionally, instead of relying on our society learning it the "hard way."

To conclude, let me reiterate that I am excited about the positive impacts that AI can have, but I also want to be clear about and mindful of the risks it gives rise to. Today, my aim is to highlight one family of such risks. I am optimistic that we can mitigate these risks, but this will require work. It cannot be left to chance. And we need to get started now.

Thank you and I am looking forward to your questions.

Acknowledgements

I am grateful for invaluable help from Sarah Cen, David Goldston, Andrew Ilyas, and Luis Videgaray.

References

[1] Sai Balasubramanian. AI offers promise and peril in tackling loneliness. *Forbes*. https://www.forbes.com/sites/saibala/2023/05/17/can-artificial-intelligence-solve-the-growing-mental-health-crisis/.

[2] Karen Hao. Deepfake porn is ruining women's lives. Now the law may finally ban it. *Technology Review*. https://www.technologyreview.com/2021/02/12/1018222/deepfake-revenge-porn-coming-ban/.

[3] Faith Karimi. 'Mom, these bad men have me': She believes scammers cloned her daughter's voice in a fake kidnapping. *CNN.com*. https://www.cnn.com/2023/04/29/us/ai-scam-calls-kidnapping-cec/index.html.

[4] John Kirchenbauer, Jonas Geiping, Yuxin Wen, Manli Shu, Khalid Saifullah, Kezhi Kong, Kasun Fernando, Aniruddha Saha, Micah Goldblum, and Tom Goldstein. On the reliability of watermarks for large language models. In *Arxiv preprint arXiv:2306.04634*, 2023.

[5] Oksana Kulyk, Nina Gerber, Annika Hilt, and Melanie Volkamer. Has the GDPR hype affected users' reaction to cookie disclaimers? *Journal of Cybersecurity*, 6(1), 12 2020.

[6] Yisroel Mirsky and Wenke Lee. The creation and detection of deepfakes. In *Arxiv preprint arXiv:2004.11138*, 2020.

[7] Johan Moreno. Canva opens up access to docs in beta, adds "magic write" generative AI copywriting tools. *Forbes*. https://www.forbes.com/sites/johanmoreno/2022/12/07/canva-opens-up-access-to-docs-in-beta-adds-magic-write-generative-ai-copywriting-tools/.

[8] Vivek H. Murthy. Our epidemic of loneliness and isolation. *The U.S. Surgeon General's Advisory*. https://www.hhs.gov/sites/default/files/surgeon-general-social-connection-advisory.pdf.

[9] Donie O'Sullivan and Jon Passantino. 'Verified' Twitter accounts share fake image of 'explosion' near Pentagon, causing confusion. *CNN.com*. https://www.cnn.com/2023/05/22/tech/twitter-fake-image-pentagon-explosion/index.html.

[10] Kevin Roose. An A.I.-generated picture won an art prize. artists aren't happy. *New York Times*. https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html.

[11] Hadi Salman, Alaa Khaddaj, Guillaume Leclerc, Andrew Ilyas, and Aleksander Mądry. Raising the cost of malicious ai-powered image editing. In *Arxiv preprint arXiv:2302.06588*, 2023.

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

[13] James Vincent. DeSantis attack ad uses fake AI images of Trump embracing Fauci. *The Verge*. https://www.theverge.com/2023/6/8/23753626/deepfake-political-attack-ad-ron-desantis-donald-trump-anthony-fauci.

[14] Yuxin Wen, John Kirchenbauer, Jonas Geiping, and Tom Goldstein. Tree-ring watermarks: Fingerprints for diffusion images that are invisible and robust. In *Arxiv preprint arXiv:2305.20030*, 2023.