S. HRG. 118–600

THE NEED FOR TRANSPARENCY IN ARTIFICIAL INTELLIGENCE

HEARING

BEFORE THE

SUBCOMMITTEE ON CONSUMER PROTECTION, PRODUCT SAFETY, AND DATA SECURITY

OF THE

COMMITTEE ON COMMERCE, SCIENCE, AND TRANSPORTATION
UNITED STATES SENATE

ONE HUNDRED EIGHTEENTH CONGRESS

FIRST SESSION

SEPTEMBER 12, 2023

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

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

U.S. GOVERNMENT PUBLISHING OFFICE
59–704 PDF WASHINGTON : 2025

SENATE COMMITTEE ON COMMERCE, SCIENCE, AND TRANSPORTATION

ONE HUNDRED EIGHTEENTH CONGRESS

FIRST SESSION

MARIA CANTWELL, Washington, *Chair*

AMY KLOBUCHAR, Minnesota	TED CRUZ, Texas, *Ranking*
BRIAN SCHATZ, Hawaii	JOHN THUNE, South Dakota
EDWARD MARKEY, Massachusetts	ROGER WICKER, Mississippi
GARY PETERS, Michigan	DEB FISCHER, Nebraska
TAMMY BALDWIN, Wisconsin	JERRY MORAN, Kansas
TAMMY DUCKWORTH, Illinois	DAN SULLIVAN, Alaska
JON TESTER, Montana	MARSHA BLACKBURN, Tennessee
KYRSTEN SINEMA, Arizona	TODD YOUNG, Indiana
JACKY ROSEN, Nevada	TED BUDD, North Carolina
BEN RAY LUJÁN, New Mexico	ERIC SCHMITT, Missouri
JOHN HICKENLOOPER, Colorado	J. D. VANCE, Ohio
RAPHAEL WARNOCK, Georgia	SHELLEY MOORE CAPITO, West Virginia
PETER WELCH, Vermont	CYNTHIA LUMMIS, Wyoming

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

SUBCOMMITTEE ON CONSUMER PROTECTION, PRODUCT SAFETY, AND DATA SECURITY

JOHN HICKENLOOPER, Colorado, *Chair*	MARSHA BLACKBURN, Tennessee, *Ranking*
AMY KLOBUCHAR, Minnesota	DEB FISCHER, Nebraska
BRIAN SCHATZ, Hawaii	JERRY MORAN, Kansas
EDWARD MARKEY, Massachusetts	DAN SULLIVAN, Alaska
TAMMY BALDWIN, Wisconsin	TODD YOUNG, Indiana
TAMMY DUCKWORTH, Illinois	TED BUDD, North Carolina
BEN RAY LUJÁN, New Mexico	CYNTHIA LUMMIS, Wyoming
PETER WELCH, Vermont	

CONTENTS

WITNESSES

APPENDIX

THE NEED FOR TRANSPARENCY IN ARTIFICIAL INTELLIGENCE

TUESDAY, SEPTEMBER 12, 2023

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

The Committee met, pursuant to notice, at 2:33 p.m., in room SR–253, Russell Senate Office Building, Hon. John Hickenlooper, Chairman of the Subcommittee, presiding.

Present: Senators Hickenlooper [presiding], Cantwell, Klobuchar, Baldwin, Luján, Welch, Blackburn, Fischer, Moran, and Young.

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

Senator HICKENLOOPER. Good afternoon. Welcome to this hearing of the Subcommittee on Consumer Protection, Product Safety and Data Security will now come to order.

While Artificial Intelligence has been part of our lives for years and years, its newer forms have now captured, it is fair to say, the world's attention. We are now far beyond the era of asking Alexa to play a song, or Siri to dial our spouse. These are the examples of Narrow-AI.

ChatGPT, the new generative AI systems can now plan a custom travel itinerary, create artwork, remix a song, and help you write computer code. So it is obvious that AI is a powerful technology that will revolutionize our economy.Just like the first car, or personal computer, AI is a transformative technology that has just both benefits and risks for consumers. That means we have to proceed with intention and care.

Our goal today is to identify how we do that. Specifically we need to begin to help Americans understand AI's capabilities and limitations, to reduce AI's potential risks, relative to consumers, and to increase the public's trust in AI systems through transparency.

The fact that we need to be careful with AI doesn't negate how important it is, or the massive potential it has to transform our lives. From helping with your tedious daily tasks, to helping doctors properly diagnose and find the right treatments for an illness, the possibilities go beyond what we can imagine today, far beyond.

But we must also confront the fact that AI can be misused by bad actors, AI can be used to make scams, fraud, and cyber attacks more harmful, and more effective. Companies developing and de-

ploying AI, we believe, have a role to build a safe, secure, and reliable system that, over time, will earn the trust of the public.

Congress will play a role by setting reasonable rules of the road to inform and protect consumers. The federal government, academia, and private sector will all need to work together to establish thoughtful AI Policy.

In April, Senator Blackburn and I sent a letter to tech companies asking how they are adopting the NIST AI Risk Management Framework. The responses showed how the framework is helping companies build accountability, transparency, and fairness into their products.

Today, Senator Thune and I sent a letter to the Office of Science and Technology Policy to stress the importance of developing Federal standards to help consumers understand and identify AI-generated content.

This is going to be more critical for building trust as AI expands into larger and larger aspects of our lives.Several other federal AI initiatives are currently underway. To name a few: The White House has convened leading tech companies, bringing people together to build a shared understanding and a voluntary commitment to build trustworthy AI systems.

NIST formed a group, a Public Generative AI Working Group to build on its AI Risk Management Framework. Also the National AI Initiative Office is coordinating a whole-of-government effort to develop AI safety and transparency with guidelines—with input from experts in civil society, in academia, and the private sector.

And we are fortunate to have two N-A-I-A-C, or NAIAC, two NAIAC members as witnesses here today. These are all encouraging steps, but it doesn't mean we have done—that we are done when it comes to making sure we have created a framework in which AI will be safe and transparent for consumers.

The AI-powered future comes with many challenges that we can already see, building a talented STEM-trained workforce, providing efficient computing power, ensuring that we protect consumer data privacy.

We know that AI trains on publicly available data, and this data can be collected from everyday consumers everywhere, in all parts of their lives. There are too many open questions about what rights people have to their own data and how it is used, which is why Congress needs to pass comprehensive data privacy protections. This will empower consumers, creators, and help us grow our modern AI-enabled economy.

This issue is complicated and it is going to require bipartisanship to acquire results. Committees across Congress are examining AI's impact on society through different lenses, each hearing is an invitation for policymakers and families at dinner tables across America to think about how AI will impact their everyday lives.

Today's discussion is that next step as we work towards building what ultimately will become, hopefully, necessarily a global consensus. This committee is well positioned to examine all of these important issues with the goal of promoting transparency, and the goal of creating an AI system that consumers will have confidence in.

I would like to welcome each of our witnesses who are joining us today: Ms. Victoria Espinel, CEO of Business Software Alliance, BSA; Dr. Ramayya Krishnan, Dean of the College of Information Systems, Carnegie Mellon University; Mr. Sam Gregory, Executive Director, WITNESS, I guess I should say, W–I-T–N-E–S-S; and Mr. Rob Strayer, Executive Vice President for Policy, Information Technology Industry Council, ITI.

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

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

Senator BLACKBURN. And thank you, Mr. Chairman. I certainly appreciate that we are having this hearing. This is kind of AI Week on the Hill, and we have a Judiciary Committee hearing going on this afternoon. And of course, we have our Member Forum, that is going to take place tomorrow and Thursday. So we are pleased to be putting attention on this.

You know, AI as you said, has been around for years, whether it is autocorrect, or auto fill, or voice assist, or facial recognition, things that people have become accustomed to using. But with ChatGPT in November, it is like people said: Wait a minute, what is this? And of course, generative AI is something that people have turned their attention to and saying: How is this happening? How is it taking place?

And Tennessee, my state, is really quite a leader in this field of AI, we have several automobile companies that are investing and innovating with AVs, we have farmers that are really leading the way in smart agriculture, and it is so interesting to hear some of their concepts. We have got thought leaders, people at the University of Tennessee, and also Oak Ridge National Lab, who are pushing boundaries on AI every single day.

With innovators like these, the future of AI can sometimes border on the unimaginable, especially as the technology continues advancing at a pace that is more rapid than any other in our recent memory.

This swift rate of advancement, however, has caused many concerns. Many of the discussions that I have heard around AI has focused on the doomsday scenarios. While it is important to prevent catastrophic events, we must also not lose sight of the transformational benefits of AI.

For instance, in addition to the examples that I have previously mentioned, AI has profound implications for the financial and health care industries, two industries that are critical to our state. That is why any regulatory action from Congress, or from Federal agencies, must balance safety with the preservation of an innovative economy.

Our adversaries, like China, are not slowing down on AI, and we cannot give them any advantage in the deployment of this emerging technology. In fact, earlier this year, Senator Ossoff and I convened a hearing on the Judiciary Subcommittee on Human Rights and the Law, where our witnesses discussed the Chinese Communist Party's interest in rolling out AI systems to enhance the regime's ability to surveil their citizens, which brings us here today.

This afternoon, we will explore ways to mitigate consumer harm, promote trust, and transparency, and identify potential risk stemming from AI technologies. And I am looking forward to hearing from each of you on these.

But before we turn to our witnesses, I wanted to remind my fellow lawmakers that amidst the hyper focus on AI, we must not forget about other issues that are as critical to U.S. technological advancement and global leadership.

First, for a decade I have worked to bring about a comprehensive data privacy law. As you mentioned, that is something that should be a first step. And I know Madam Chairman is well aware, and joins me in wanting to see a Federal standard. And it is vital that my colleagues keep in mind the need for that Federal privacy standard, as we look at AI. In our judiciary hearings that we have had on AI, everybody mentioned the need to have that so people can protect their data.

It is virtually impossible to talk about these new and advanced systems without a real discussion about how online consumers will be able to protect what I term, "their virtual you", which is their presence online.

Second, as AI systems require more and more computing power, the need for high performance in quantum computing will become vital. This is why I have already introduced two bipartisan bills on this topic, and I encourage this committee to move on reauthorizing the National Quantum Initiative Act. We need to do that this year.

So thank you to each of you, for the leadership you have on the issue for being here as our witnesses, and thank you for the hearing.

Senator HICKENLOOPER. Thank you, Senator Blackburn.

I am now going to turn it over to the Chair of the Commerce Committee, who has a long, a long history, probably more direct history with AI than any other senator. Senator Cantwell from Washington.

STATEMENT OF HON. MARIA CANTWELL,
U.S. SENATOR FROM WASHINGTON

The CHAIR. Thank you, Mr. Chairman. And thank you to yourself; and to Senator Blackburn at the Subcommittee level for holding this important hearing. I think we are demonstrating that just as AI needs to be open and transparent, we are going to have an open and transparent process as we consider legislation in this area.

And I want to thank Senator Blackburn for her comments about privacy because I do think these things go hand in hand, having good, strong privacy protections certainly prevents the kind of abuse or misuse of information that could cause substantial harm to individuals.

And I thank the witnesses for being here today to help us in this discussion.

I recently was informed about a situation in my state that I found quite alarming. A family in Pierce County, Washington, received a phone call. A scammer used AI to spoof the voice of their daughter, telling them that she had been in a car accident and that a man was threatening to harm her if they didn't wire $10,000. So

I can't imagine what this deepfake meant to that family or the concerns that they have.

And a recent deepfake image claimed a bombing occurred at the Pentagon, and that fake image sparked a dip in the stock market.

DARPA is leading the way on important developments to approach detecting AI-generated media. And I plan to introduce legislation in this area.

I think that AI, as was discussed by my two colleagues, has amazing potential. I held an AI Summit in my state and saw some of those amazing technologies already being pushed by the Allen Brain Institute and some of their early technologies; certainly helping in things like climate, and farming, and detecting illegal activities, and helping us move forward in important areas of research.

So we know that we have choices here. We know we want to continue to empower consumers and make sure that we are stopping the fraudsters. And we want to make sure that any misuse of AI that we are stopping that, whatever we can do to make sure that we are protecting American's privacies.

So I hope that today's hearing will give us some ideas about how to drive innovation and maintain U.S. leadership in this very important security-related technology, and the issues of global competitiveness, that we talk and discover ideas, about deepfakes and potential national security issues, the framework for legislation, protect online privacy, and combat discrimination.

I know that we need to grow education, in general, in our workforce. And the information age has already put great transformations in place. The jobs of tomorrow are here today, but the skill levels for people to do them are not.

We know that we need to invest more from the CHIPS and Science Act in skilling a workforce for tomorrow. That was before AI. With AI, there is an accelerant on that. And that is why I believe we need something as grand as the GI Bill was after World War II in empowering Americans for new opportunities in this area.

So I look forward to hearing the comments from our witnesses.

And thank you again Mr. Chairman, for holding this very important hearing about the potential and challenges facing us. But clearly, we need an open and transparent system, just as we did for the internet, so that innovation can flourish. Thank you.

Senator HICKENLOOPER. Thank you, Madam Chair. I appreciate your being here and helping create this hearing. Now we will go and we will have the opening statements from each of the witnesses.

Ms. Espinel.

STATEMENT OF VICTORIA ESPINEL, CHIEF EXECUTIVE OFFICER, BSA | THE SOFTWARE ALLIANCE

Ms. ESPINEL. Good afternoon, Chair Hickenlooper, Ranking Member Blackburn, Chair Cantwell, and Members of the Subcommittee.

My name is Victoria Espinel, and I am the CEO of BSA | The Software Alliance.

BSA is the advocate for the global enterprise software industry. BSA members are at the forefront of developing cutting-edge serv-

ices including AI, and their products are used by businesses across every sector of the economy. I commend the Subcommittee for convening today's hearing, and I thank you for the opportunity to testify.

There are two things that need to be done. Companies that develop and use AI must act responsibly to identify and address risks, and Congress needs to establish thoughtful, effective rules that protect consumers, and promote responsible innovation.

AI has real world benefits. Think about extreme weather events, hurricanes, wildfires, tornadoes, that have affected many states this year, as we know, there are families wondering whether the eye of a hurricane will hit their hometown, and whether they will be safe if they stay, or if they should pack up and go. How will they know whether they should leave? And if they do, which nearby destination is the safest to ride out the storm?

AI is helping to provide these answers. With AI weather forecasters are better able to predict extreme weather events, helping people prepare before disaster strikes. And what happens to those families who are in the storm's path? How do they get food in the aftermath of a storm? How do rescue workers know that they need help?

AI is helping relief workers anticipate where medical equipment, food, water, and supplies are most needed in response to natural disasters.

In the face of extreme danger, AI's predictions can save lives. More needs to be done, however, so that we can see greater benefits. And with thoughtful rules in place, innovation in AI will continue to advance, and the responsible use of artificial intelligence will serve society.

There has been a wave of attention on AI since ChatGPT launched publicly nine months ago. But this committee began studying the issue in a thoughtful manner years earlier. Nearly six years ago, I testified here about the building blocks of machine learning and artificial intelligence. Chair Cantwell and Senator Young introduced one of the first AI bills in 2017. We also appreciate the committee's recent work to establish the National Artificial Intelligence Initiative, and your request to BSA about how our member companies are using the NIST, AI Risk Management Framework to responsibly develop and use AI.

The pace of AI development and use has increased significantly since 2017. As with any new technology there are legitimate concerns that need to be addressed, including the risk of bias and discrimination.

This committee is well placed to move legislation that sets rules around AI. The U.S. economy will benefit from responsible and broad-based AI adoption. An important part of facilitating that adoption is passing a strong national law. The countries that best support responsible AI innovation will see the greatest benefits of economic and job growth in the coming years.

Moreover, other countries are moving quickly on regulations that affect U.S. companies. The U.S. should be part of shaping the global approach to responsible AI. The window for the U.S. to lead those conversations, globally, is rapidly closing.

This is what we think legislation should do: It should focus on high risk uses of AI, like those that decide whether a person can get a job, a home, health care. It should require companies to have risk management programs. It should require companies to conduct impact assessments, and it should require companies to publicly certify that they have met those requirements.

This will include some concrete steps. I have set these out in more detail in my written testimony, and I hope we have a chance to discuss those.

It is important that legislation reflect different roles. Some companies develop AI, some companies use AI, our companies do both, and both roles have to be covered. Legislation should set distinct obligations for developers and users because each will know different things about the AI system in question, and each will be able to take different actions to identify and mitigate risks.

So my message to Congress is simple. Do not wait. AI legislation can build on work by governmental organizations, industry, and civil society. These steps provide a collective basis for action. You can develop and pass AI legislation now that creates meaningful rules to reduce risks and promote innovation. We are ready to help you do so.

Thank you for the opportunity to testify. And I look forward to your questions.

[The prepared statement of Ms. Espinel follows:]

PREPARED STATEMENT OF VICTORIA ESPINEL, CHIEF EXECUTIVE OFFICER, BSA | THE SOFTWARE ALLIANCE

Good afternoon Chairman Hickenlooper, Ranking Member Blackburn, and members of the Subcommittee. My name is Victoria Espinel, and I am the CEO of BSA | The Software Alliance.[1]

BSA is the leading advocate for the global enterprise software industry.[2] Our members are at the forefront of developing cutting-edge services—including AI—and their products are used by businesses across every sector of the economy. I commend the Subcommittee for convening today's hearing and thank you for the opportunity to testify. I also appreciate this Committee's longstanding focus on AI, including your efforts to establish the National Artificial Intelligence Initiative and your outreach to BSA to learn about how our companies are implementing the AI Risk Management Framework released earlier this year by the National Institute of Standards and Technology.

Nearly six years ago, I testified before this Committee at a hearing focused on the building blocks of machine learning and artificial intelligence.[3] Chair Cantwell and Senator Young also introduced one of the first AI bills that year. Since then, the building blocks we discussed in 2017 have come together at a rapid pace. Traditional benchmarks for measuring how AI performs tasks like recognizing and

[1] I am a member of the National AI Advisory Committee, but I am testifying in my capacity as CEO of BSA.

[2] BSA's members include: Adobe, Alteryx, Atlassian, Autodesk, Bentley Systems, Box, Cisco, CNC/Mastercam, Databricks, DocuSign, Dropbox, Elastic, Graphisoft, IBM, Informatica, Juniper Networks, Kyndryl, MathWorks, Microsoft, Okta, Oracle, Palo Alto Networks, Prokon, PTC, Rubrik, Salesforce, SAP, ServiceNow, Shopify Inc., Siemens Industry Software Inc., Splunk, Trend Micro, Trimble Solutions Corporation, TriNet, Twilio, Unity Technologies, Inc., Workday, Zendesk, and Zoom Video Communications, Inc.

[3] Testimony of Victoria Espinel, Hearing on Digital Decision-Making: The Building Blocks of Machine Learning and Artificial Intelligence, Before the Senate Committee on Commerce, Science, & Transportation Subcommittee on Communications, Technology, Innovation, and the Internet (Dec. 12, 2017), available at *https://www.bsa.org/files/policy-filings/12122017BSAAI Testimony.pdf*.

classifying images or understanding text are becoming obsolete, as researchers launch new methods to measure progress.[4]

As I said then, AI is a foundational technology that drives products and services that people use every day. It also raises important policy issues, which are core to our work at BSA. We undertook a year-long project to work with member companies to develop the BSA Framework to Build Trust in AI,[5] which was released in 2021 and is designed to help organizations mitigate the potential for unintended bias in AI systems. Built on a vast body of research, the BSA Framework sets out a lifecycle-based approach for performing impact assessments to identify risks and highlights best practices for mitigating those risks.

Best practice documents, like BSA's Framework, have moved the policy debate forward, but I am here today to say that they are not enough, and legislation is needed. Thoughtful AI legislation will benefit the U.S. economy by creating new guardrails that build trust in the use of AI technologies. It will protect consumers by ensuring AI developers and deployers take required steps to mitigate risks; and it will set the U.S. as a leader in the global debate about the right way to regulate AI. Fortunately, Congress can leverage tools that already exist to create legislation that requires companies to identify and mitigate risks associated with high-risk AI systems.

I. Congress Should Act Now to Adopt Meaningful AI Legislation

Congress should not wait to enact legislation that creates new obligations for companies that develop and use AI in high-risk ways.

Legislation will not only protect consumers from real risks of harm, but will also create trust in AI technologies that will benefit the economy broadly. Consumers and businesses already rely on a wide range of AI-powered services, but they will only continue to adopt new AI technologies if they trust that those products and services are developed and deployed responsibly. Because companies of all sizes and in all industry sectors can benefit from AI, thoughtful AI legislation is important to promoting the United States economy. Countries that support the broad adoption of AI will see the greatest growth in jobs and economic prosperity.

To enact legislation, Congress should take advantage of the considerable work that governmental organizations, civil society advocates, and industry groups have put into identifying the risks of using AI in different contexts and the concrete steps organizations can take to mitigate those risks. Although these proposals have important differences, they collectively form a basis for action. For example, there are fundamental objectives on which everyone should agree: AI, in any form, should not be used to commit illegal acts. It should not be used to compromise privacy, facilitate cyberattacks, exacerbate discrimination, or create physical harm. AI that is developed and deployed responsibly, that improves our lives and makes us safer, should flourish. But Congress should go farther, to create new rules for companies that develop and deploy high-risk AI systems.

By enacting legislation, Congress will also ensure the United States is not just a leader in developing AI technology but is a leading voice in the global debate about regulating AI. The window to lead conversations about AI regulation is rapidly closing, as other governments are moving to shape the rules that will govern AI's future. By the end of this year, the European Union is expected to finalize its AI Act, which is on pace to be the most comprehensive AI law enacted. Japan is leading a G7 effort to establish common standards for AI governance. In November, the United Kingdom will host a global AI summit, focused on the safe and responsible deployment of AI. These governments are not alone. Australia, Brazil, Canada, China, Korea, Singapore, and Thailand are among the long list of countries that are examining the right policy approaches for addressing AI risks.[6]

---

[4] AI Index Steering Committee at the Institute for Human-Centered AI at Stanford University, The AI Index 2023 Annual Report (April 2023), page 114, available at *https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf* (AI Index 2023).

[5] *See* Confronting Bias: BSA's Framework to Build Trust in AI (June 2021), available at *https://www.bsa.org/reports/confronting-bias-bsas-framework-to-build-trust-in-ai.* BSA has testified before the United States Congress and the European Parliament on the Framework.

[6] *See* Australia: BSA Comments on Supporting Safe and Responsible AI Innovation in Australia (July 26, 2023), available at *https://www.bsa.org/files/policy-filings/07262023safeai.pdf;* Brazil: BSA Recommendations to the Jurists Committee on AI Regulation in Brazil (May 30, 2022), available at *https://www.bsa.org/files/policy-filingsen05302022airegbrazil.pdf;* Korea: BSA Comments on Bill on Fostering AI Industry and Securing Trust (Feb. 13, 2023), available at *https://www.bsa.org/files/policy-filings/en02132023kraitrust.pdf;* Singapore: BSA Comments on Public Consultation on Proposed Advisory Guidelines on Use of Personal Data in AI Recommendation and Decision Systems (Aug. 30, 2023), available at *https://www.bsa.org/files/policy-filings/08302023bsaaiguidelines.pdf;* Thailand: BSA Comments on Draft Bill on Promotion

The United States has recognized the enormous benefits of working with other countries on AI governance issues, and is already participating in a range of global efforts. We support the United States' active involvement in these fora, including the US–EU Trade and Technology Council (TTC), the Global Partnership on AI (GPAI), the Organisation for Economic Co-Operation and Development (OECD), and the G7 Hiroshima AI project. But the U.S. voice in those efforts will be stronger if the United States adopts national AI legislation that creates clear guardrails for how companies develop and deploy high-risk AI systems.

My message to Congress is: Do not wait. You can adopt AI legislation now that creates meaningful rules to reduce risks and promote innovation. We urge you to do so.

II. The U.S. Economy Will Benefit from Legislation that Builds Trust in AI

Businesses of all sizes across all industries are looking for ways to adopt AI tools to grow. Indeed, consumers and businesses today already rely on a wide range of services powered by enterprise AI—and the AI our companies create is all around us. While these AI systems may not get the most attention, they are increasingly integrated in the economy and everyday life. For example:

- ***Consumers.*** As consumers, AI-powered services remind us when we forget to attach a document to an e-mail, or prompt us to open a particular document so that we can pick up where we left off. We expect to auto-complete forms, chat with virtual assistants that are available even when customer service representatives are not, and to use AI-powered apps to identify and reach personal financial goals, or to predict how a sports player will perform in an upcoming game.
- ***Businesses.*** Businesses across every industry sector are integrating AI into their products and services, including retail stores that use AI to predict when demand for a product will surge so that they can keep shelves stocked, warehouses that rely on AI-powered logistics planning to minimize supply chain shortages, and manufacturers that use AI to detect safety concerns on factory floors. Across industry sectors, businesses of all sizes also use AI to identify and manage common documents, detect fraudulent transactions, and guard against cybersecurity threats.[7]

The economic benefits of AI are not limited to one industry sector or one business model. Instead, the promise that AI may one day impact every industry is quickly turning into a commercial reality and driving digital transformation across sectors. Airlines now use AI systems to more efficiently clean planes between flights; farmers use AI to analyze large amounts of weather information to maximize their harvest; manufacturers use AI to test new prototypes, and construction companies build AI-generated "digital twins" of real-life cities to understand the impacts of a proposed design.

Enterprise software companies are at the leading edge of this transformation, creating the products and services relied on by other companies.[8] I have included an extensive list of examples from BSA's members in an annex to this testimony, but want to highlight a handful of ways that companies in all industries are using enterprise-powered AI:

- In healthcare, a large pharmacy chain uses an AI-powered platform to forecast demand and redistribute medications across thousands of store locations and to deliver near real-time insights and recommendations for pharmacists to provide more personalized advice to patients. This helps managers understand the supply chain, store labor and productivity, patient vaccine scheduling, and prescription pickup processes.
- In manufacturing, a car maker used AI-based generative technology to redesign a seat bracket, which secures seat belt fasteners to seats and seats to floors, that is 40 percent lighter and 20 percent stronger than the previous iteration. Changes like these can help reduce the amount of material needed to build a car and make vehicles more fuel efficient.

---

and Support of National AI Innovation (Aug. 18, 2023), available at *https://www.bsa.org/files/policy-filings/08182023bsanatlai.pdf.*

[7] For more examples of everyday uses of AI, *see* BSA, Everyday AI for Consumers, available at *https://www.bsa.org/files/policy-filings/08012023aiconsumers.pdf,* and BSA, Everyday AI for Businesses, available at *https://www.bsa.org/files/policy-filings/08012023aibusiness.pdf.*

[8] BSA's policy and educational resources on AI are available at *https://ai.bsa.org/*. For additional information about AI's adoption across industry sectors *see* BSA, Artificial Intelligence in Every Sector, available at *https://www.bsa.org/files/policy-filings/06132022bsaaieverysector.pdf.*

- In agriculture, the research division of an enterprise software provider partnered with a climate risk company to develop AI models capable of providing more accurate long-range weather predictions. Traditional weather forecasting methods can provide accurate predictions for a seven-day window. By leveraging AI, the researchers are developing new forecasting models to provide accurate predictions of weather trends two-to six-weeks out from a given date. By providing reliable extended forecasts, these tools will help water managers predict snowpack and water availability for irrigation, hydropower, and other critical agricultural and environmental uses.

The importance of AI to the U.S. economy is inescapable. Innovations in electricity and personal computers created investment booms of as much as 2 percent of U.S. GDP as those technologies were adopted into the broader economy.[9] Economic forecasts estimate that AI could have an even bigger impact on GDP. By 2025, investment in AI is expected to approach $100 billion in the United States and $200 billion globally.[10] Generative AI alone could add up to $4.4 trillion of value to the global economy every year.[11]

Why does this matter for policymakers?

To realize these economic benefits, consumers and businesses must trust that AI is developed and deployed responsibly. While the adoption of AI can unquestionably be a force for good, it can also create real risks if not developed and deployed responsibly. Setting thoughtful rules for AI is therefore central to the vitality of our economy. Industries of all kinds and businesses of all sizes are looking for ways to use AI to grow, but they will only adopt AI-powered products and services that they trust. Countries that best facilitate responsible and broad-based AI adoption will see the greatest economic and job growth in the coming years.

III. Congress Should Enact Legislation That Builds Trust in AI

Now is the time for Congress to adopt thoughtful legislation that addresses known risks of AI systems.

There will continue to be a range of significant and evolving policy debates around AI. But there's no need to wait to pass legislation that creates meaningful guardrails against the AI risks that exist today. You can—and should—build on existing regulatory efforts by setting rules across the economy to address concerns about the potential for bias when AI systems are used in high-risk ways.

A. Legislation Should Create New Obligations for High-Risk Uses of AI

AI legislation can build on work that's already been done.

To anticipate and address potential harms, developers of AI used in high-risk situations should understand the practical implications of its use, including through the use of impact assessments. Those who train AI systems for high-risk uses, or are deploying AI systems in high-risk contexts, should also be required to understand how the tools they create or deploy might result in unintended outcomes and act to mitigate those risks.

We urge Congress to adopt legislation that creates meaningful guardrails for high-risk uses of AI. That legislation should require companies to:

(1) establish risk management programs to identify and mitigate risks across AI systems;

(2) conduct annual impact assessments for high-risk uses of AI, and

(3) publicly certify that they have met these requirements.

Congress can build on tools that already exist today—and require companies to adopt those tools to identify and mitigate risks associated with high-risk AI systems. My testimony focuses on two of those tools: risk management programs and impact assessments.

B. Risk Management Programs

The goal of risk management is establishing repeatable processes for identifying and mitigating potential risks that can arise throughout the lifecycle of an AI system. Risk management is particularly important in contexts like AI, privacy, and

---

[9] Goldman Sachs, AI Investment Forecast to Approach $200 Billion Globally By 2025 (Aug. 1, 2023), available at *https://www.goldmansachs.com/intelligence/pages/ai-investment-forecast-to-approach-200-billion-globally-by-2025.html.*

[10] *Id.*

[11] McKinsey & Company, The Economic Potential of Generative AI (June 2023), available at *https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier?gclid=Cj0KCQjwl8anBhCFARIsAKbbpyTeLnz0c6i4X2UTnmWdO1KGQnE1mUR8ErJSrM0eMnWDxgfaZukt_L0aAqP7EALw_wcB#/.*

cybersecurity, where the combination of quickly evolving technologies and highly dynamic threat landscapes can render traditional approaches ineffective. Rather than evaluating a product or service against a static set of requirements that can rapidly become outdated, risk management programs integrate compliance responsibilities into the development process to help identify and mitigate risks throughout a product or service's lifecycle.

Risk management programs have two key elements: (1) a governance framework to support the organizations' risk management functions, and (2) a scalable process for performing impact assessments that identify and mitigate risks of an AI system. The governance framework is critical because it promotes collaboration between an organization's development team and its compliance team at key points in the design, development, and deployment of an AI system.

One way for companies to establish risk management programs is by using the AI Risk Management Framework (AI RMF), which was released earlier this year by the National Institute of Standards and Technology (NIST).[12] Creating the AI RMF was a significant achievement and builds on NIST's work creating frameworks for managing cybersecurity and privacy risks. For example, the NIST Cybersecurity Framework (CSF) is widely used by private and public-sector organizations worldwide; since 2017, it has been mandatory for Federal agencies to use the CSF to improve their cybersecurity risk management programs.[13] Like the CSF, the AI RMF is a voluntary tool. It helps organizations incorporate trustworthiness considerations into the design, development, use, and evaluation of AI products.

The AI RMF can help companies establish risk management programs that guard against a range of potential AI-related harms. It focuses on corporate practices around four functions: govern, map, measure, and manage. The AI RMF is designed to be usable by large and small organizations alike and can be applied in varied circumstances, such as for procurement purposes, for organizations with existing governance programs, and for organizations just beginning to think about risk management. It also identifies several indicia of trustworthy AI, which include privacy, explainability, and fairness, and incorporates an assessment of these characteristics as part of the measure function. Importantly, the AI RMF acknowledges that tradeoffs among trustworthiness characteristics may exist. For example, the use of privacy-enhancing technologies may decrease accuracy, which could affect decisions about fairness in certain domains, and there may also be tradeoffs between interpretability and privacy.

The AI RMF encourages:

- Consultation with diverse stakeholders;
- Establishing processes to identify, assess, and mitigate risks;
- Defining individual roles and responsibilities to people throughout an organization;
- Identifying metrics for evaluation;
- Evaluating fairness and bias;
- Maintaining post-deployment feedback mechanisms; and
- Establishing incident response plans.

Ultimately, effective AI risk management programs should be underpinned by a governance framework that establishes the policies, processes, and personnel that will be used to identify, mitigate, and document risks throughout the system's lifecycle. The purpose of such a governance framework is to promote understanding across organizational units—including product development, compliance, marketing, sales, and senior management—about each entity's role and responsibilities for promoting effective risk management during the design, development, and deployment of AI systems.

C. Impact Assessments

Impact assessments have three purposes: (1) identifying potential risks that an AI system may pose, (2) quantifying the degree of potential harms the system could generate, and (3) documenting steps taken to mitigate those risks.[14] As noted ear-

---

[12] NIST AI Risk Management Framework, available at *https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100–1.pdf.*

[13] *See* NIST, Cybersecurity Framework, Questions and Answers, (discussing Federal agency use of the NIST CSF), available at *https://www.nist.gov/cyberframework/frequently-asked-questions/framework-basics#agency.*

[14] *See* BSA, Impact Assessments: A Key Part of AI Accountability, available at *https://www.bsa.org/files/policy-filings/08012023impactassess.pdf.*

lier, performing impact assessments is a key part of creating a meaningful risk management program.

Impact assessments are already widely used in a range of other fields, including data protection, as an accountability mechanism that demonstrates a system has been designed in a manner that accounts for the potential risks it may pose to the public. Because impact assessments already exist today, they can be readily adapted to help companies identify and mitigate AI-related risks. For example, three state privacy laws already require companies to conduct impact assessment for specific activities, including processing sensitive personal data, engaging in targeted advertising, or selling personal data; seven more state privacy laws will soon do so.[15] Globally, privacy and data protection laws worldwide use impact assessments as a tool for improving accountability.[16]

The use of impact assessments can be readily expanded to require assessments for high-risk AI systems. Conducting impact assessments can help companies identify and mitigate risks, including risks of unintended bias. In our view, when AI is used in ways that could adversely impact civil rights or access to important life opportunities, the public should be assured that such systems have been thoroughly vetted and will be continuously monitored to account for the risks associated with unintended bias. Requiring impact assessments for companies that develop and deploy these high-risk systems is an important way to do that.

D. Legislation Should Focus on High-Risk Uses and Create Role-Appropriate Requirements.

Legislation that requires risk management programs and impact assessment will create new safeguards for high-risk AI systems. Any legislation incorporating these requirements should:

- *Focus on high-risk AI uses.* Legislation should place guardrails around high-risk uses of AI, namely AI systems that determine an individual's eligibility for housing, employment, credit, education, access to physical places of public accommodation, healthcare, or insurance. These systems have the potential to affect important life opportunities—and are a key area for policymakers to address. In contrast, many everyday uses of AI present few risks to individuals and create significant benefits, like helping organize digital files, auto-populate common forms for later human review, or predict the fonts to use in a template document. Indeed, Americans have reported being most excited about the potential for AI to perform household chores and conduct repetitive workplace tasks, while expressing significant concerns about the use of AI to make important life decisions for people.[17] Legislation should address these high-risk uses.
- *Recognize the different roles of companies that develop AI systems and companies that deploy AI systems.* Legislation should recognize the different roles of companies that create and use AI systems. Developers are the companies that design, code, or produce an AI system, such as a software company that develops an AI system for speech recognition. In contrast, deployers are the companies that use an AI system, such as a bank that uses an AI system to make loan determinations. Legislation should recognize the different roles of developers and deployers, because these two types of companies will have access to

[15] Colorado, Connecticut, and Virginia already impose these requirements. See Colorado Privacy Act, Colo. Rev. Stat. Tit. 6, Art. 1, Pt. 13 §§ 6–1–1301–6–1–1313; Connecticut Data Privacy Act Conn. Gen. Stat. Tit. 42, Ch. 743jj, Sec. 42–515–525; Virginia Consumer Data Protection Act; Va. Code Tit. 59.1, Ch. 53, § 59.1–575–585. Recently passed state privacy laws in Florida, Indiana, Montana, Oregon, Tennessee, and Texas will also require impact assessments for certain activities.

[16] Data protection impact assessments are an established accountability tool under privacy laws worldwide. Under the European Union's General Data Protection Regulation (GDPR), impact assessments are required for activities likely to result in a "high risk" to the rights and freedoms of individuals. Brazil's General Data Protection Law (LGPD) also allows the national data protection agency to require a company to prepare a data protection assessment, and in Singapore organizations are required to conduct impact assessments if they process data for certain purposes. There is also significant regulatory guidance on how organizations can conduct impact assessments to identify and mitigate privacy risks. *See, e.g.,* Office of the Australian Information Commissioner, Guide to Undertaking Privacy Impact Assessments (Sept. 2, 2021), available at *https://www.oaic.gov.au/privacy/privacy-guidance-for-organisations-and-government-agencies/privacy-impact-assessments/guide-to-undertaking-privacy-impact-assessments;* European Data Protection Board, Guidelines on Data Protection Impact Assessment (Oct. 13, 2017), available at *https://ec.europa.eu/newsroom/article29/items/611236;* Singapore Personal Data Protection Commission, Guide to Data Protection Impact Assessments, (Sept. 14, 2021) available at *https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Other-Guides/DPIA/Guide-to-Data-Protection-Impact-Assessments-14-Sep-2021.pdf.*

[17] AI Index 2023, page 329.

different types of information and will be able to take different actions to mitigate risks.

For example, the developer of an AI system is well positioned to describe features of the data used to train that system, but it generally will not have insight to how the system is used after it is purchased by another company and deployed. A company that deploys an AI system is well positioned to understand how the system is actually being used, what type of human oversight is in place, and whether there are complaints about how the system works in practice. Legislation should recognize these different roles, so that the appropriate company in the real-world AI supply chain can identify and mitigate risks. For the same reasons, this kind of distinction is considered best practice in privacy and security legislation around the world.[18]

Legislation that leverages these existing tools can crate meaningful safeguards for the development and deployment of AI in high-risk uses.

E. Benefits of Legislation

If Congress passes a law that requires risk management programs and impact assessments, it can require companies to identify and mitigate across the lifecycle of an AI system. These requirements will have real benefits to consumers, by making companies evaluate and address potential risks that can arise across the lifecycle of an AI system. These risks can arise in contexts including:

- *Problem Formulation.* The initial step in building an AI system is often referred to as "problem formulation." It involves the identification and specification of the "problem" the system is intended to address, an initial mapping of how the model will achieve that objective, and the identification of a "target variable" the system will be used to predict. Because many AI systems are designed to make predictions about attributes that are not directly measurable, data scientists must often identify variables that can be used as proxies for the quality or outcome it is intended to predict.

  While the use of proxy target variables can be entirely reasonable, the assumptions underlying the choice of proxies must be closely scrutinized to ensure that it does not introduce unintended bias to the system. The risk that can arise in problem formulation is exemplified by a study of a widely used healthcare algorithm that hospitals rely on to identify patients in need of urgent care. The research team concluded that the algorithm was systematically assigning lower risk scores to black patients compared to similarly sick white counterparts because it relied on data about historical healthcare costs as proxy for predicting a patient's future healthcare needs. Unfortunately, because black patients have historically had less access to healthcare, the reliance of spending data painted an inaccurate picture and led to dangerously biased outcomes.[19]
- *Training Data.* The data used to train an AI system is a second major vector for bias. If the data used to train a system is misrepresentative of the population in which it will be used, there is a risk the system will perform less effectively on communities that may be underrepresented in the training data. Likewise, reliance on data that itself may be the product of institutional or historical biases can entrench those inequities in an AI model.
- *Labeling.* The process of "labelling" training data can also introduce bias. Many AI systems require training data to be "labeled" so that the learning algorithm can identify patterns and correlations that can be used to classify future data inputs. Because the process of labeling the data can involve subjective decisions, there is the potential for introducing unintended bias into the training data.
- *Deployment.* Even a system thoroughly vetted during development can begin to exhibit bias after it is deployed. AI systems are trained on data that represents a static moment in time and that filters out "noise" that could undermine the model's ability to make consistent and accurate predictions. Upon deployment in the real world, AI systems inevitably encounter conditions that differ from those in the development and testing environment. Because the real world changes over time, the snapshot in time that a model represents may naturally become less accurate. If the input data for a deployed AI system differs materially from its training data, there is a risk that the system could "drift" and that the performance of the model could be undermined. For instance, if an AI sys-

[18] *See* BSA, AI Developers and Deployers: An Important Distinction, available at *https://www.bsa.org/files/policy-filings/03162023aidevdep.pdf.*

[19] Heidi Ledford, Millions of Black People Affected by Racial Bias in Health-Care Algorithms, Nature (Oct. 24, 2019), available at *https://www.nature.com/articles/d41586-019-03228-6.*

tem is designed (and tested) for use in a specific country, the system may not perform well if it is deployed in a country with radically different demographics. Bias can also arise if an AI system is deployed into an environment that differs significantly from the conditions for which it was designed or for purposes that are inconsistent with its intended use.

Congress should act now to require companies that develop or deploy high-risk AI to adopt risk management programs and to conduct impact assessments. These requirements will protect consumers and create meaningful rules that reduce risks and promote innovation.

* * *

We appreciate Congress's leadership on the important policy issues raised by AI. We are ready to help as you craft and pass legislation. Thank you and I look forward to your questions.

ANNEX: AI IN EVERY SECTOR

Improving Healthcare and Quality of Life

The rapid digitalization of health information has created tremendous opportunities for AI to transform how clinicians care for patients, how consumers manage their health, and how researchers discover breakthroughs in the treatment and prevention of diseases.

Helping Pharmacies Redistribute Medication and Provide Personalized Advice to Patients

Walgreens uses the Databricks Lakehouse platform to run an intelligent data platform incorporating AI to forecast demand and redistribute medications across Walgreens' network of nearly 9,000 pharmacies, while delivering near real-time insights and recommendations for pharmacists to help provide more personalized advice to patients. This integrated AI-driven platform allows Walgreens' different data teams to work better together, create smarter algorithms and generate new types of reporting to help managers understand the supply chain, store labor and productivity, patient vaccine scheduling, and prescription pickup processes.

Advancing Accessibility

For people with visual impairments, AI is turning the visual world into an audible experience. Microsoft's Seeing AI app helps people who are blind or visually impaired recognize objects, people, and text via a phone or tablet's camera and describes what it recognizes to the user. With this new layer of information, users can navigate the world more independently.

Strengthening Security

Although data security is core to the management of most organizations, cyber threats continue to evolve at a breakneck pace. AI helps organizations stay a step ahead of hackers by predicting potential attacks, mitigating attacks in real-time, managing access to resources, and encrypting sensitive data.

Enabling Fast Action Against Security Threats

Palo Alto Networks' AI-driven Security Operations Center automation engine, XSIAM, is delivering never-before-seen cybersecurity outcomes. The company's own instance of this tool ingests 36 billion events every day from across all network layers and attack surfaces and triages just 8 of those for human analysis. This empowers their most precious resources—people—to focus on the most sophisticated attacks that uniquely require human analysis. Importantly, this AI-driven tool has reduced overall Mean Time to Detection (MTTD) to 10 seconds and Mean Time to Response (MTTR) to one minute for high priority incidents. This more resilient and automated cyber future would not be possible without AI.

Protecting Business Transactions

Splunk is helping financial institutions to leverage AI and data analytics to strengthen their cybersecurity and their ability to serve customers. For example, consumer report and risk scoring provider TransUnion uses data analytics and machine learning capabilities provided by Splunk to monitor customer traffic and transactions. TransUnion monitors and manages customer traffic to its website and detects when unusual activity takes place so it can alert customers about security concerns and ensure seamless customer experiences.

Building 21st Century Infrastructure

Whether it's creating smarter and safer cities by integrating sensors in bridges and highways to monitor their safety or increasing efficiency by cutting travel time and fuel expenses, AI plays an instrumental role in creating an infrastructure designed for the 21st century.

Optimizing Manufacturing

Generative design tools can optimize the manufacturing process to reduce waste and improve products. Autodesk teamed up with Michigan-based foundry Aristo Cast to develop an ultralightweight aircraft seat frame. The team used generative design, 3D printing, lattice optimization, and investment casting to ultimately create a seat frame that weighs 56 percent less than typical current models. For a 615-seat Airbus A380 plane, that would mean saving $100,000 in fuel per year, as well as more than 140,000 fewer tons of carbon in the atmosphere.

Streamlining Building Projects

Companies are using AI to streamline the building design and construction processes. Bentley Systems has teamed with Hyundai engineering on an AI system that automates design processes for steel and concrete structures, reducing the time needed to create designs and the cost of building a structure.

Monitoring Vehicle Fleets

Oracle's anomaly detection software uses AI to monitor the operation of complex systems and detect potentially concerning incidents. Transportation and logistics company SS Global LLC uses Oracle's software to monitor their fleet of vehicles and get alerts when there are early signs of potential safety issues. By detecting the early onset of tire baldness and air leaks, the system helps SS Global perform predictive maintenance that keeps their fleet safer and more efficient.

Creating New Ways to Learn

AI applications are enabling personalized learning resources for every stage of life, including adaptive learning programs, digital tutoring, curriculum recommendations, and more. There are more digital resources available to instructors and students than ever before, and AI is affording them the ability to access relevant tools quickly and easily.

Enriching Math Education

Educators are using IBM's Teacher Advisor With Watson AI to access the math resources they need in seconds, including proven lesson plans, activities, standards information, and teaching strategies for students with varying degrees of preparation and ability. This can save valuable time for teachers throughout the school year.

Tailoring Workplace Learning

Employers are using Workday Learning, an application that uses machine learning to personalize workplace learning for individuals, to recommend professional development content and courses based on employee position, tenure at the company, interactions with the content, and other factors. This helps companies adjust learning strategies and programming to ensure employees learn new skills, continue to grow in their roles, and prepare for what's ahead.

Enhancing the Customer Experience

For businesses with large customer bases that are processing a high volume of purchases—such as banks, restaurant chains, and large retailers—analyzing the massive amount of data collected every day is impossible without the computing and predictive power of AI. By using machine learning tools, businesses across a wide range of industries can analyze customer preferences and their own business performance to improve end-user experiences and increase efficiencies. Software also helps businesses generate optimal product designs by using data to produce and analyze far more iterations than humans alone could create.

Customizing Care Experiences

Powered by Salesforce AI technology, Eli Lilly has reimagined patient care with its Patient Connect Platform app. The app helps customers learn to use products, access information about their medications, and record how well they are feeling. The desktop and mobile apps also allow patients to consult with a healthcare concierge—a specialist who provides one-on-one support to guide patients toward beneficial health outcomes.

Improving Customer Service Experience

Zendesk is using AI to improve the customer service experience for both customers and the agents that interact with them. Using Zendesk's intelligent triage functionality, a company can automatically detect a customer's intent (for example, whether a customer is making a return or checking on shipping status), the language the customer is using, and the customer's overall sentiment so that the inquiry can be quickly routed to the best agent for the job. Several of Zendesk's business-to-consumer customers are using this Zendesk AI feature to automatically classify and route incoming tickets to the right agents at the right time, which has resulted in higher customer satisfaction and more one-touch tickets.

Scaling Community Impact

Twilio provides AI chatbot services to help businesses interact with customers. The United Way Worldwide worked with Twilio to help scale and route inbound calls and texts to more than 200 agencies nationwide that use their 211 system to help people locate essential needs like housing, financial assistance, food, childcare, transportation, and more. Using the AI-assisted interactive voice response menu built with Twilio Autopilot, the United Way and Twilio built a system that enables a caller to access a single 1–800 number or be transferred by their local 211 to access assistance. The result is a centralized system that efficiently reduces the call volume nationwide but increases the time staffers are able to devote to mission critical calls.

Improving Business Operations

AI is helping to streamline business operations and increase productivity.

Enhancing Business Functions

SAP provides chatbot solutions that are seamlessly integrated into other business functions, giving customers, partners, and employees a bird's-eye view of business operations. For example, SAP provides software services to Hewlett Packard Enterprise Company, including an AI-based chatbot system that can reference serial numbers, packing slips, and shipment dates drawn from cloud services, thereby getting the right information to the right people at the right time.

Improving Contract Analysis

DocuSign has been helping organizations use AI-based technologies including natural language processing and rules-based logic to manage and analyze agreements for several years now. Using AI-powered contract analysis can increase productivity in the contract process by helping to speed up contract reviews, increase contract visibility, and identify opportunities and risks.

Empowering Creativity

AI and machine learning within Adobe's Creative Cloud tools help artists, photographers, designers, and content creators around the world handle the time-consuming aspects of their work that can easily be automated, so they have more time to be creative. From removing unwanted objects like mics and logos from videos in Adobe After Effects, to colorizing black-and-white photos in just a few clicks in Adobe Photoshop, to painting with digital brushes that look, feel, and act like the real thing in Adobe Fresco, Adobe's AI and machine learning features empower creators to focus their energy on what they love—ideating, experimenting, and creating.

Helping in Times of Crisis

In times of humanitarian crises, fast response is essential. Researchers are developing ways to use AI to help first responders in the critical hours and days after a natural disaster, and to track pathogens that could lead to outbreaks of disease and mitigate the spread.

Navigating the COVID–19 Pandemic

Siemens' Dynamic VAV Optimization (DVO) is a software solution for building management systems that uses machine learning and AI to configure HVAC controls according to a building's priorities, whether that's minimizing virus transmission or minimizing energy consumption. In direct response to the challenges of the pandemic, DVO was launched with a new operating Defense Mode in late 2020 to reduce the risk of viral spread in indoor spaces. DVO adjusts ventilation, temperature, and humidity conditions to minimize risk of viral spread indoors while also maximizing energy efficiency.

Enriching Our Lives

Leveling Up Gaming and Entertainment

AI can be used to create sophisticated 3–D environments and train autonomous characters in our favorite games and movies. Unity's AI products are used to develop video games, animations, and other detailed virtual environments. By training computer-based characters in Unity's software, game designers can create more realistic environments that capture a player's imagination and enhance the gaming experience.

Senator HICKENLOOPER. Thank you, Ms. Espinel.
Dr. Krishnan.

STATEMENT OF DR. RAMAYYA KRISHNAN, W. W. COOPER AND RUTH F. COOPER PROFESSOR OF MANAGEMENT SCIENCE AND INFORMATION SYSTEMS; DEAN, HEINZ COLLEGE OF INFORMATION SYSTEMS AND PUBLIC POLICY; FOUNDING FACULTY DIRECTOR, THE BLOCK CENTER FOR TECHNOLOGY AND SOCIETY, CARNEGIE MELLON UNIVERSITY

Dr. KRISHNAN. Good afternoon. Chair Cantwell, Chairman Hickenlooper, Ranking Member Blackburn, and Members of the Committee, I am grateful for this opportunity to testify today. My name is Ramayya Krishnan. I am the Dean of the Heinz College of Information Systems and Public Policy. I have served as President of INFORMS, a Global Operations Research and Analytics Society, and my perspective is shaped by my own work in this area, as well as my leadership of The Block Center for Technology and Society, a university-wide initiative that studies responsible use of AI and the future of work.

I am a member of the NAIAC, but I not here representing the NAIAC.

You have already heard about the vast number of applications and the potential for AI. I would like to highlight that in addition to all of that, AI has the capacity to enable breakthroughs in science and drug discovery that unlock solutions that are currently intractable—currently intractable problems in human health, and beyond.

These are among the many important socially and economically beneficial uses of the technology. As AI technologies are considered for use in high-stakes applications, such as autonomous vehicles, health care, recruiting, criminal justice, the unwillingness of leading vendors to disclose the attributes and provenance of the data that they have used to train and tune their models, and the processes they have employed for model training and alignment to minimize the risk of toxic or harmful responses, needs to be urgently addressed.

This lack of transparency creates threats to privacy, security, and the uncompensated use of intellectual property to copyrighted content. In addition to the harms caused to individuals and community due to bias and unreliable performance, there is a need for greater accountability and transparency.

In my testimony I would like to propose four decisive recommendations. The first, is on promoting responsible AI, Congress should require all federal agencies to use the NIST AI RMF in design, development, procurement, use and management of their AI use cases. The NIST AI RMF was developed with multiple stake-

holder inputs, and establishing it as a standard that will have numerous benefits at home and abroad.

My next recommendation is a two-part recommendation. The first relates to advancing data transparency in the AI pipeline. The AI pipeline consists of training data, models, and applications, and data transparency, Congress should require standardized documentation, and like audited financial statements, they should be verifiable by a trusted third party, like an auditor, about the AI training data.

The metaphor to think about is to think of these as akin to nutrition labels, so it is clear what went into producing the model. The details about what these should contain, but at the minimum it should document the sources and rights that the model developers have, to be able to use the data that they did, and the structural analysis that they have done to check for biases, and the like. And perhaps even for adversarial attacks against this data.

The second part of this recommendation is promoting model validation and evaluation of the AI system. Congress should direct NIST to develop standards for high-stakes domains, such as health care, recruiting, criminal justice, and require model validation report for AI systems deployed in high-stakes applications.

The metaphor here is to think of this being akin to an underwriter's lab-like report that will objectively assess the risk and performance of an AI system in these high-stakes applications.

The third recommendation is on content transparency, it is about content labeling and detection. The main idea here is that, as we have just heard, the generative AI has created—increased the capacity to create multi-modal content, audio, video, text, that can be—that is indistinguishable from human created output, and currently there is no standardized way to label the content as being AI generated, or human generated.

There are consortia like C2PA that are coming around with standards, but we need standards here, and Congress should require all AI models, open source, or otherwise, that produce content to be able to label their content with watermarking technology, and provide a tool to detect that label.

While the usual concern about labeling is with regard to consumers, this is going to be equally important for model developers to know if the data that they are using in their models, is human-generated or AI-generated.

The last recommendation is about investing in a trust infrastructure for AI, much like we did in the late-1980s, when we stood up a trust infrastructure in response to cybersecurity attacks, the Morris worm in 1988, and we set up the CERT, the Computer Emergency Response Team. We need to do something similar for AI.

Congress should stand up a capability which could be done relatively quickly using the capacity of FFRDCs, and NIST in the SEI MITRE, and other agencies, Federal Government agencies that connect vendors, catalog incidents, record vulnerabilities, and test and verify models, and disseminate best practices.

This will go a long way towards improving our trust capability, especially since the technology is moving so quickly, that we will always have to meet this quick-response capability.

Finally, in closing, the success of these recommendations will in part rest on a comprehensive approach to enhance AI skills across K through 12, and community colleges, as well as policies and strategies, like wage insurance to address the impact of AI.

Thank you for this opportunity to testify to the committee.

[The prepared statement of Dr. Krishnan follows:]

PREPARED STATEMENT OF DR. RAMAYYA KRISHNAN, W. W. COOPER AND RUTH F. COOPER PROFESSOR OF MANAGEMENT SCIENCE AND INFORMATION SYSTEMS; DEAN, HEINZ COLLEGE OF INFORMATION SYSTEMS AND PUBLIC POLICY; FOUNDING FACULTY DIRECTOR, THE BLOCK CENTER FOR TECHNOLOGY AND SOCIETY, CARNEGIE MELLON UNIVERSITY

Chair Cantwell, Ranking Member Cruz, Subcommittee Chair Hickenlooper, Ranking Member Blackburn and Members of the Committee, I am grateful for the opportunity to testify today. My name is Ramayya Krishnan and I serve as Dean of the Heinz College of Information Systems and Public Policy, a multidisciplinary academic unit that spans information technology, data analytics and public policy at Carnegie Mellon University. My perspective is shaped by my work over several decades on the use of data analytics and advanced model-based approaches to support decision making in consequential applications and by my role as the faculty director of the Block Center for Technology and Society, a university wide initiative that studies the responsible use of AI and its consequences for the future of work. I am a member of the National AI Advisory Committee (NAIAC). However, I am here in my own capacity and not representing the NAIAC.

Artificial intelligence (AI) is a transformative technology. Its application to create personalized tutors (See *https://youtu.be/yEgHrxvLsz0*), create operational and clinical decision support tools in health care (*https://www.nature.com/articles/s41586–023–06160-y*), promote health literacy among citizens *https://www.cmu.edu/news/stories/archives/2023/august/revolutionizing-health-care-harnessing-artificial-intelligence-for-better-patient-care*), and enable breakthroughs in science and drug discovery that will unlock solutions to currently intractable problems in human health and beyond are among the many economically and societally beneficial uses of the technology. And likely many of us have used an AI chatbot like chatGPT, a generative AI technology, and seen both its immense potential and its failures.

As AI technologies are considered for use in high stakes applications such as in health care, recruiting and criminal justice, the unwillingness of the leading vendors to disclose the attributes and provenance of the data they have used to train and tune their models, and the processes they have employed for model training and "alignment" to minimize the risk of toxic or harmful responses needs to be urgently addressed (*https://arxiv.org/pdf/1901.10002.pdf*). This lack of transparency creates threats to privacy, security, and uncompensated use of intellectual property and copyrighted content in addition to harms caused to individuals and communities, due to biased and unreliable performance. There is a need for greater accountability and transparency in the development and deployment of AI to spur its responsible adoption and use.

In my testimony, I propose four decisive recommendations for Congress to consider to address these challenges. Investing in these recommendations will provide near term impact on trusted adoption of AI and, when combined with a focused research program, will ensure U.S. leadership in responsible and trustworthy AI.

The recommendations include foundational actions to advance broad based adoption of Responsible AI practices as well as measures to accelerate the adoption of practices and technologies to mitigate the threat of deep fakes and protect the rights of creators. My final recommendation is to create an infrastructure for AI trust by establishing a capability to monitor and respond to AI vulnerabilities and failures and support the development and dissemination of solutions and best practices. These measures will need to be closely aligned with a focused research program that advances the development of tools for watermarking and labeling as well as research into measurement, metrics and evaluation of reliability and quality of the AI supply chain. A detailed policy memo co-authored by my colleagues and myself from Carnegie Mellon on accountable AI is available at *https://www.cmu.edu/block-center/responsible-ai/cmu_blockcenter_rai-memo_final.pdf*.

1. Promoting Responsible AI

Congress should require all Federal agencies to use the NIST (National Institute of Standards and Technology) AI Risk Management Framework during the design,

development, procurement, use, and management of their AI use cases. This will promote responsible adoption and deployment of AI in government and more broadly in society. Investing in workshops such as the NIST-Carnegie Mellon AI RMF workshop which convened academics and industry representatives from sectors such as banking, health care and consulting to discuss the gaps that need to be addressed to operationalize responsible AI in the economy will be particularly valuable. The NIST AI RMF was developed with multiple stakeholder inputs and establishing it as a standard will have numerous benefits at home and abroad.

2. Promoting Greater AI Transparency

a. Content transparency: Content Labeling and Detection:

A significant concern with the advent of generative AI is the ease with which multi-modal content (audio, video text) can be created that is indistinguishable from human created content. Multiple examples document why this is a problem. Students submitting AI-produced content in lieu of their own work is an academic integrity issue and hurts learning outcomes. Audio and video deep fakes raise concerns from multiple standpoints—from affecting the economic outcomes and reputations of well-known artists to concerns for human rights and democracy.

Currently, there is no standardized way to label content as AI generated and no standardized tool that can use such labels to help consumers recognize AI generated content. Recent commitments by major vendors to develop watermarks for AI generated content and the emergence of content provenance standards (*e.g.,* C2PA) is a step in the right direction. While proposals exist for audio and visual content, watermarking and provenance for AI generated text remains a challenge. As with all security technologies, watermarking will need to stay one step ahead of attempts to defeat it (*https://legaljournal.princeton.edu/the-high-stakes-of-deepfakes-the-growing-necessity-of-federal-legislation-to-regulate-this-rapidly-evolving-technology/* and *https://gjia.georgetown.edu/2023/05/24/should-the-united-states-or-the-european-union-follow-chinas-lead-and-require-watermarks-for-generative-ai/*).

While the usual concern about this issue has been from the point of view of the consumer, this inability to distinguish AI generated content from human produced content and knowledge of its provenance is relevant to model developers as well. Since Internet content is used at scale to train models and as AI produced content proliferates on the internet, model developers will need the capacity to differentiate AI produced content from human produced content since this has implications for model performance.

Congress should require all AI models (open source and closed source models) that produce content to label their content with watermarking and provenance technology and provide a tool to detect the label.

b. Advancing Transparency in the AI pipeline for high stakes applications

The AI pipeline or value chain (Hosanagar and Krishnan, 2023) consists of training data, models and applications. It is this pipeline that is used to create AI systems in high stakes applications such as in autonomous vehicles, health care, recruiting, and criminal justice. The leading vendors of closed source models do not disclose the attributes and provenance of the data they have used to train and tune their models, and the processes they have employed for model training and "alignment" to minimize the risk of toxic or harmful responses. When these AI pipelines are used in high stakes applications, greater transparency around how AI is integrated into the broader system is needed. We can learn from prior work in accountable AI as well as in modeling reliability of societal and engineered systems to address these transparency questions. In the following, I will highlight key recommendations that pertain to data transparency and AI model validation and evaluation. These measures are vital to address critical challenges created by Large Language Models. It will also be vital in enabling U.S. companies to remain globally competitive as international standards are developed.

- *Advancing Data Transparency*
- Model developers need to document the rights they have to work with the data they are using to train the model. This documentation should also provide information about the source of the data, whether it was public or private, etc.
- Model developers should respect the right of data owners to opt out of data crawling (robots.txt file) and also provide data owners the opportunity to opt out of the use of their already collected data in model training or tuning.
- Model developers need to document the standards that were used in bias assessment and demonstrate the analysis that was conducted to assess structural bias in the data.

Congress should require standardized documentation and, like audited financial statements, they should be verifiable by a trusted third party (e.g., an auditor). The metaphor is to think of these as akin to "nutrition labels." so it is clear what went into producing the model.

- *Promoting Model Validation and Evaluation of the AI system*
- Develop clear standards for articulating intended use cases and metrics for reliability and utility so that users can have clear expectations of performance under well-defined conditions. *Congress should direct NIST to develop standards for these high stakes applications such as health care, recruiting, and criminal justice.*
- Define AI sandboxes and test data sets, evaluation frameworks, measurement and metrics, and continuous monitoring standards based on the assessed risk of the application space or use case.
- Require the auditor to use these standards and validation infrastructure to evaluate the AI system and provide the required assurance prior to deployment.

Congress should require a model validation report for AI systems deployed in high stakes applications. The metaphor is to think of this as akin to an "underwriters lab" that objectively assesses the risk and performance of an AI system.

c. Investing in a trust infrastructure for AI (AI CERT), or ALRT (AI lead response team)

AI is developing rapidly. Even with the proposed transparency measures, there will be a need to respond rapidly to newly discovered AI vulnerabilities, exploits and failures. The safety and reliability of the AI ecosystem is a necessary condition to engender trust and spur widespread adoption and deployment of AI. While AI Incident databases (*https://incidentdatabase.ai/*) from the Responsible AI Collaborative, Project Atlas from MITRE (see *https://atlas.mitre.org/*) and the recently organized DEFCON red teaming event and *voluntary commitments* are important steps forward, an institutional solution is required.

The proposed solution, an ALRT, would connect vendors, AI system deployers and users. It would catalog incidents, record vulnerabilities, test and verify models, and recommend solutions and share best practices to minimize systemic risks (*https://www.forbes.com/sites/rosecelestin/2023/08/23/the-ai-financial-crisis-theory-demystified-how-to-create-resilient-global-ecosystems/?sh=27282b4d51ce*) as well as harm stemming from vulnerability exploits. This is modeled after the computer emergency response team (CERT) that the U.S. Government stood up in response to cyber security vulnerabilities and threats in late 1980s. The following capabilities are required to serve the needs of CERT for AI)

- Catalog incidents, record vulnerabilities, recommend solutions and, share best practices to minimize risks
- Coordinate commercial and public government focus with the need to rapidly respond to national security challenges (*e.g.,* chemical/bio weapons *https://arxiv.org/ftp/arxiv/papers/2306/2306.03809.pdf*). Be able to respond to threats that affect .com, .gov and .mil domains. In effect, combine open, restricted and classified work
- Possess deep technical capabilities spanning core AI and computing and an understanding of how the core AI is operationalized to meet application needs
- Maintain domain knowledge connected to applications
- Convene industry, government and academic partners around core tech AI technology as well as operationalization of AI

Congress should stand up these capabilities quickly via existing FFRDCs and harness the strengths at NIST and other Federal agencies. implementation of these recommendations will have an immediate impact on trusted adoption of AI (e.g., standing up ALRT). Combining implementation with investments in a focused program of research on Responsible AI and AI transparency will ensure U.S. leadership in trustworthy AI.

Finally, in closing, the success of these recommendations will in part rest on comprehensive strategies that enhance AI skills across the continuum from K–12 (https://www.tutors.plus/) and community college {https://sailplatform.org/) education to new tools, strategies and policies (https://tinyurl.com/2528ke6z) to support workers in virtually all industries to adapt to the impact of AI. Thank you for this opportunity to testify to your committee.

Senator HICKENLOOPER. Thank you, Dr. Krishnan.

Mr. Gregory.

STATEMENT OF SAM GREGORY, EXECUTIVE DIRECTOR, WITNESS

Mr. GREGORY. Chairman Hickenlooper, Ranking Member Blackburn, and Members of the Subcommittee; I am Sam Gregory, Executive Director of WITNESS, a human rights organization.

Since 2018, WITNESS has led a global effort, Prepare, Don't Panic, to inclusively prepare society for deepfakes in synthetic media technologies, and more recently generative AI. Now, the moment to act has come.

Increased volumes of easily made realistic synthetic photos, video, and audio, more targeted and personalized is a paradigm shift. Alongside creative and commercial benefits from generative AI, there are already harms in the U.S. and globally, with disproportionate impacts on at-risk groups.

Women are targeted with nonconsensual sexual images; simulated audio scams are proliferating, as are AI generated child sexual abuse images. Claims of AI generation are used to dismiss critical human rights and journalistic content that is real. Text-to-image tools perpetuate discriminatory patterns, and creatives see their work used for training AI models without consent.

As you develop legislation, first consult broadly with the communities most impacted by AI in the U.S. and globally. Given differential impacts, a resilient approach grounded in civil and human rights, will best future proof legislative responses to AI.

Elections are poised to be deeply influenced by generative AI, and recent polls find the American public is fearful of its impact. Underresourced newsrooms and community leaders are under pressure, and do not have access to reliable tools that can detect deepfakes and AI-manipulated content.

It is also unreasonable to expect consumers to spot deceptive yet realistic imagery and voices. Guidance to look for the six-fingered hand, or spot virtual errors in the Pope in the puffer jacket, do not help in the long run.

I believe that Congress should think about AI governance as a pipeline of responsibility that is distributed across technology actors from the foundation models to those designing and deploying software and apps, and to platforms that disseminate content. This should be supported by testing, auditing, transparency, and pre-post-impact assessment.

Within that, solutions that help show the provenance of AI, and if desired, human-generated content can use this pipeline responsibility to bring transparency to consumers. This may take the shape of direct visible, or indirect machine readable disclosure around how media is made, created, and/or distributed. Approaches like this were included in the White House Voluntary Commitments, and in the EU's AI Act.

As I note in my written testimony, visible watermarks and direct disclosure have use cases. For example, potentially, in marking election-related materials, but they are easily removed, and not nuanced for the complex production era that we are about to enter.

Invisible watermarks at the pixel and dataset level are another option.

Cryptographically signed metadata like the C2PA standard show the creation, production, and distribution process over time, which is important—is important as we increasingly intermingle AI and human generated.

Approaches like this also allow creators to indicate how and if their content may be used for training AI models. These disclosure approaches are not a punitive measure to single out AI content, nor indicate deception. This provenance data only provides additional information signals, but does not provide any truth to the content.

To safeguard constitutional and human rights approaches to provenance could meet some core criteria. They should first protect privacy. Personally identifiable information should not be required. Knowing the how of AI-based production is key, but does not require correlating the identify of who made the content, or instructed the tool. Allowing redaction is also key, particularly when combining AI-based media with other human-generated media.

Secondly, opt-in. While disclose indicating content was AI-generated could be a legal requirement in certain cases, it shouldn't be a requirement for all provenance tools, especially those for non-AI content, which should always be opt-in.

And thirdly, standards should be developed attentive to potential authoritarian misuse in a global context.

Another response complementary to content provenance, is after-the-fact detection for content believed to be AI-generated. From witnesses' experience, the skills and tools to detect AI-generated media remain unavailable to the people who need them most. Journalists, rights defenders, and election officials domestically and globally; it remains critical to support Federal research and investment in this area to close this detection access and equity gap.

To conclude, I encourage you to incorporate broad consultation with groups disproportionately experiencing existing harms and AI impacts into upcoming legislative approaches, to go beyond risk-based and voluntary approaches, and support a rights-based framework for action.

And finally, to support research and legislation on standardized watermarking and provenance that takes into account global implications, and centers privacy and accessibility.

Thank you for the opportunity to testify today.

[The prepared statement of Mr. Gregory follows:]

TESTIMONY OF SAM GREGORY, EXECUTIVE DIRECTOR, WITNESS

Chairman Hickenlooper, Ranking Member Blackburn and members of the Senate Commerce Subcommittee on Consumer Protection, Product Safety and Data Security, thank you for the opportunity to testify today about transparency in AI.

I am Sam Gregory, Executive Director of WITNESS, a human rights organization.[1] Since 2018, WITNESS has led a global effort, Prepare, Don't Panic, to understand how deepfake and synthetic media technologies, and more recently large language models (LLMs) and generative AI, are impacting consumers, citizens and communities in the U.S. and globally, and to prepare accordingly.[2] These efforts have included contribution to technical standards development,[3] engagement on de-

[1] WITNESS *https://www.witness.org/*

[2] For our work on generative AI and deepfakes see: *https://www.gen-ai.witness.org/*

[3] Jacobo Castellanos, *WITNESS and the C2PA Harms and Misuse Assessment Process,* WITNESS, December 2021, *https://blog.witness.org/2021/12/witness-and-the-c2pa-harms-and-misuse-assessment-process/*

tection and authenticity approaches that can support consumer literacy,[4] analysis and real-time response to contemporary usages,[5] research,[6] and consultative work with rights defenders, journalists, content creators, technologists and other members of civil society.[7] Our experience is further informed by three decades of experience helping communities, citizens, journalists and human rights defenders create trustworthy photos and videos related to critical societal issues and protect themselves against the misuse of their content.

Today, I will have a particular focus on how to optimize the benefits, and minimize the harms and risks from multimodal audiovisual generative AI. These tools, with their potential to create realistic image, audio and video simulations at scale, as well as personalized content, will have far-reaching implications for consumers, creative production and generally, our trust in the information we see and hear.

Executive Summary

My organization, WITNESS, has for a number of years promoted a perspective of *Prepare, Don't Panic* in relation to deepfakes and generative AI. But the moment to act and prepare society for audiovisual generative AI and its impacts has come.

Transparency around AI's role in the production of information that consumers and citizens engage with is a critical area. In this testimony, I will focus on questions of disclosure and provenance in audiovisual content, and how these relate to the responsibility of actors in the AI pipeline. WITNESS, in consultation with global experts and communities affected by technology development, focuses on three overarching principles to guide the assessment of the opportunities and risks that generative AI brings to society.

1. Place firm responsibility on stakeholders across the AI, technology and information pipeline.
2. Center those who are protecting human rights and democracy, and communities most impacted by AI, domestically and globally, in the development of solutions.
3. Embed human rights standards and a rights-based approach in the response to AI.

With these principles in mind, U.S. policy makers and legislators have a range of options to promote transparency in AI and protect consumers and their data:

1. Ensure broad consultations with communities impacted by AI when developing solutions to watermarking, provenance and disclosure, and in broader processes of transparency common to all AI systems–including documentation, third-party auditing, pre-release testing, evaluation, and human rights impact monitoring.
2. Push for a rights-based approach to transparency in AI, that promotes standardized systems for disclosing when content has been made with AI, while supporting opt-in solutions that track the provenance of non-synthetic media.
3. Prohibit personal data from being included by default in any approaches to provenance, disclosure and watermarking, for all types of content.

---

[4] WITNESS Media Lab, *How do we work together to detect AI-generated media? https://lab.witness.org/projects/osint-digital-forensics/*

[5] Nilesh Christopher, *An Indian politician says scandalous audio clips are AI deepfakes: We had them tested,* Rest of World, July 2023, *https://restofworld.org/2023/indian-politician-leaked-audio-ai-deepfake/*

[6] Gabriela Ivens and Sam Gregory, *Ticks or It Didn't Happen: Confronting Key Dilemmas in Authenticity Infrastructure for Multimedia,* WITNESS, December 2019, *https://lab.witness.org/ticks-or-it-didnt-happen/*

[7] Raquel Vazquez Llorente, Jacobo Castellanos and Nkem Agunwa, *Fortifying the Truth in the Age of Synthetic Media and Generative AI.* WITNESS, June 2023, *https://blog.witness.org/2023/05/generative-ai-africa/;* Sam Gregory, *Deepfakes, misinformation and disinformation and authenticity infrastructure responses: Impacts on frontline witnessing, distant witnessing, and civic journalism.* Journalism, December 2021, *https://www.researchgate.net/publication/356976532_Deepfakes_misinformation_and_disinformation_and_authenticity_infrastructure_responses_Impacts_on_frontline_witnessing_distant_witnessing_and_civic_journalism.* Also, see: *Deepfakes: Prepare Now (Perspectives from Brazil),* WITNESS, 2019, *https://lab.witness.org/brazil-deepfakes-prepare-now/; Deepfakes: Prepare Now (Perspectives from South and Southeast Asia),* WITNESS, 2020 *https://lab.witness.org/asia-deepfakes-prepare-now/;* Corin Faife, *What We Learned from the Pretoria Deepfakes Workshop,* WITNESS, 2020, *https://blog.witness.org/2020/02/report-pretoria-deepfakes-workshop/;* Corin Faife, *How Can U.S. Activists Confront Deepfakes and Disinformation?* WITNESS, 2020, *https://blog.witness.org/2020/12/usa-activists-disinformation-deepfakes/*

4. Enact comprehensive data privacy legislation, as well as integrate data privacy rights into broader AI legislation that also includes solutions for opting out of models' datasets.
5. Support research and investment in technologies that can detect AI manipulation and generation and are accessible domestically and globally, as well as consumer-facing tools that provide information on content provenance.

The domestic and global context of audiovisual generative AI and deepfakes

While there are creative and commercial benefits to generative AI and synthetic media, these tools are also already connected to a range of harms to U.S. consumers and global users. Chatbots provide incorrect, factual-appearing information. Audio scams using simulated audio are proliferating. Non-consensual sexual images are used to target private citizens and public figures, particularly women, and AI-generated child sexual abuse images are increasing. Claims of AI-generation are used to dismiss verifiable content. Text-to-image tools perpetuate existing patterns of bias or discriminatory representation present in their training data. Creatives and artists have had their production incorporated into training for AI models without consent, and no-one has access to reliable ways to opt their images out of these training data sets.[8]

In the area that I focus on, audiovisual generative AI and deepfakes, research indicates that humans do trust the realism cues of audio and video,[9] cannot identify machine-generated speech cloning accurately,[10] do not recognize simulated human faces,[11] do not fare well spotting face-swapped faces,[12] and retain false memories of deepfakes.[13] In the direct experience of my own organization in analyzing high-profile suspected deepfakes encountered globally, it is challenging to support rapid, high-quality media forensics analysis; detection resources are not widely available to the media or the public; and the gap between analysis and timely public understanding is wide and can be easily exploited by malicious actors.[14]

These AI tools create *accidental harms* when they don't work as promised or anticipated including when they 'hallucinate' information. Deceptive information is a feature, not a bug of systems. Consumers also face *misuse harms when* generative AI tools work as intended to, but are exploited deliberately for criminal or deceptive purposes, such as cloning voices for scams.

There are *supply chain harms* derived from representational biases that are a reflection of both developers' choices and prejudices embedded in the training data,

---

[8] Rhiannan Williams, Melissa Heikkilä, You need to talk to your kid about AI. Here are 6 things you should say, MIT Tech Review, September 2023, *https://www.technologyreview.com/2023/09/05/1079009/you-need-to-talk-to-your-kid-about-ai-here-are-6-things-you-should-say/;* Matt O'Brien, Chatbots sometimes make things up. Is AI's hallucination problem fixable?, AP, August 2023, *https://apnews.com/article/artificial-intelligence-hallucination-chatbots-chatgpt-falsehoods-ac4672c5b06e6f91050aa46ee731bcf4;* FTC Consumer Alert, Scammers use AI to enhance their family emergency schemes, March 2023, *https://consumer.ftc.gov/consumer-alerts/2023/03/scammers-use-ai-enhance-their-family-emergency-schemes;* Benj Edwards, *AI-generated child sex imagery has every U.S. attorney general calling for action,* Ars Technica, September 2023, *https://arstechnica.com/information-technology/2023/09/ai-generated-child-sex-imagery-has-every-us-attorney-general-calling-for-action/;* Nilesh Christopher, *ibid;* Rida Qadri, Renee Shelby, Cynthia L. Bennett and Emily Denoton, *AI's Regimes of Representation: A Community-Centered Study of Text-to-Image Models in South Asia,* 2023, *https://arxiv.org/abs/2305.11844* ; Harry H. Jiang, Lauren Brown, Jessica Cheng, Mehtab Khan, Abhishek Gupta, Deja Workman, Alex Hanna, Johnathan Flowers, and Timnit Gebru, *AI Art and its Impact on Artists,* August 2023, *https://dl.acm.org/doi/fullHtml/10.1145/3600211.3604681;*

[9] Steven J, Frenda, Eric D. Knowles, William Saletan, Elizabeth F Loftus, *False memories of fabricated political events,* Journal of Experimental Social Psychology, 2013,*https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2201941*

[10] Hibaq Farah, *Humans can detect deepfake speech only 73 percent of the time, study finds,* The Guardian, August 2023, *https://www.theguardian.com/technology/2023/aug/02/humans-can-detect-deepfake-speech-only-73-of-the-time-study-finds*

[11] Sophie J. Nightingale and Hany Farid, *AI-synthesized faces are indistinguishable from real faces and more trustworthy,* PNAS, February 2022, *https://www.pnas.org/doi/10.1073/pnas.2120481119*

[12] Nils C. Köbis, Barbora Doležalová and Ivan Soraperra, *Fooled twice: People cannot detect deepfakes but think they can,* IScience, November 2021, *https://www.sciencedirect.com/science/article/pii/S2589004221013353*

[13] Nadine Liv, Dov Greenbaum, *Deep Fakes and Memory Malleability: False Memories in the Service of Fake News,* AJOB Neuroscience, March 2020, *https://www.tandfonline.com/doi/abs/10.1080/21507740.2020.1740351?journalCode=uabn20*

[14] Sam Gregory, *Pre-Empting a Crisis: Deepfake Detection Skills + Global Access to Media Forensics Tools,* WITNESS, *https://blog.witness.org/2021/07/deepfake-detection-skills-tools-access/ ;* Nilesh Christopher, *ibid.*

as well as harms that come from the inappropriate incorporation of personal data, creative production or intellectual property into the development processes of AI.

Finally, given the lack of public understanding of AI, the rapidly increasing verisimilitude of outputs, and the absence of robust transparency and accountability, the combination of poorly functioning and misused technology brings *structural harms*—in this case undermining broader trust in information and public processes.[15]

High-risk usages of synthetic media and generative AI may not be easily defined and will depend on context, and the potential for differential impact. Risk-based approaches to regulation transfer a lot of responsibility to the private sector and may result in regulatory gaps impacting those consumers already most at risk. A rights-based approach—not only a risk-based approach—that is grounded in U.S. Constitutional values is key to protecting the interests of consumers and citizens in the US. Existing legal frameworks on data protection and sector specific regulations for health or financial markets, for instance, provide a basis for action, as do the protections in the White House AI Bill of Rights in relation to broader AI issues, Executive Order 13960 on Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government, and the Organization for Economic Co-operation and Development's (OECD's) 2019 Recommendation on Artificial Intelligence that the U.S. adopted.

As I highlight later, comprehensive data privacy legislation and if needed AI-specific regulation that incorporates strong privacy protections would provide a core bedrock for addressing both strong implementation of safeguards for consumers such as transparency, as well as supporting future developments in AI technologies.

Synthetic media tools are now able to produce images of real-life events and realistic audio of individuals with limited input data, and at scale. Increased volume of easily made, realistic synthetic photos, audio and eventually audio, of specific real individuals and contexts is a paradigm shift. In the future, this will include more accurate, targeted and interactive personalization for a given context, individual consumer, specific user or audience in existing social media contexts, as well as emerging formats for communications such as VR and AR.[16] Generative AI tools are increasingly multimodal, with text, image, video, audio and code functioning interchangeably as input or output.

It is unreasonable to expect consumers and citizens to be able to 'spot' deceptive and realistic imagery and voices. As the Federal Trade Commission (FTC) has already noted,[17] most of the challenges and risks with generative AI cannot be addressed by the consumer acting alone.

Similarly, responses to the risks of these tools cannot be adequately addressed by regulatory agencies or laws without a pipeline of responsibility across foundation models, developers and deployers of AI models.

Single technical solutions will not be sufficient either. In the case of audiovisual generative AI, deepfakes and synthetic media, technical approaches to detection will need to be combined with privacy-protecting, accessible watermarking and opt-in provenance approaches, and with mandatory processes of documentation and transparency for foundation models, pre-release testing, third-party auditing, and pre/post-release human rights impact assessments.

Harms and risks of deepfakes, generative AI and synthetic media identified by WITNESS

Through the past five years of WITNESS consultations, civil society leaders have consistently identified a set of existing harms and potential threats from synthetic media and deepfakes. As tools have become more accessible and personalizable, and easier to use, a higher number of people have had the ability to engage with them. They have been able to imagine—or experience—how these technologies could impact their lives.

The main overarching concern echoes across countries: threats from synthetic media will disproportionately impact those who are already at risk, because of their

---

[15] Matt Davies and Michael Birtwistle, *Seizing the 'AI moment': making a success of the AI Safety Summit,* Ada Lovelace Institute, September 2023, *https://www.adalovelaceinstitute.org/blog/ai-safety-summit/;* for additional discussion of evaluating impacts, Irene Solaiman *et al., Evaluating the Social Impact of Generative AI Systems in Systems and Society,* June 2023, *https://arxiv.org/abs/2306.05949*

[16] Eric Horvitz, *On the Horizon: Interactive and Compositional Deepfakes,* 2022, *https://arxiv.org/abs/2209.01714;* Thor Benson, *This Disinformation is Just for You,* WIRED, August 2023, *https://www.wired.com/story/generative-ai-custom-disinformation/*

[17] FTC Business Blog, *Chatbots, deepfakes, and voice clones: AI deception for sale,* March 2023, *https://www.ftc.gov/business-guidance/blog/2023/03/chatbots-deepfakes-voice-clones-ai-deception-sale*

ethnicity, gender, sexual orientation, profession, or belonging to a social group. Women particularly already face widespread threats from non-consensual sexual images that do not require high-quality or complex production to be harmful. Many marginalized and vulnerable populations have already been affected by the existing AI-driven dynamics of the information ecosystem. They have experienced AI and other forms of technology that have brought differential and/or disparate impact to them. This reflects both the biases in these tools (*e.g.,* representational bias), as well as their use and misuse to disproportionately target these populations.

Elections in the coming year are poised to be deeply influenced by the malicious or deceptive use of generative AI. We hear how the fear of synthetic media, combined with the confusion about its capabilities and the lack of knowledge to detect AI-manipulation, are misused to dismiss authentic information with claims it is falsified. This is so-called plausible deniability or the "liar's dividend".[18] In our work analyzing claims of deepfakes, incidents of the liar's dividend are highly prevalent.

Similarly, these tools could be used by foreign governments to close civil society space by, for instance, incorporating them into patterns of criminalization and harassment of journalists and human rights defenders, and disinformation targeting their activities and those of political opponents at home and abroad. The potential threats brought by synthetic media and generative AI have motivated governments to enact laws suppressing free expression and dissent, posing a threat to the principles of free expression, civic debate and information sharing. Proposed rule-making and legislation on generative AI and deepfakes in China is indicative of this trend.[19]

Lastly, pressures to understand complex synthetic content, and claims that content is synthesized, place additional strain on already under-resourced local and national newsrooms and community leaders responsible for verifying digital content. With hyperbolic rhetoric as well as the realities of advances in generative AI undermining trust in content we encounter, human rights defenders, journalists and civil society actors will be among the most impacted by generative AI.

These technologies need to be developed, deployed, or regulated with an in-depth understanding of a range of other local and national contexts. The voices of those impacted by AI, need to be central to the discussion and prioritization of solutions.[20] Yet, emerging technologies are designed and deployed without the input of those most impacted, ignoring the threats and risks these technologies bring to communities already at a disadvantage.

Why consumer-facing transparency matters in an information environment with more complex creative and communicative production

Media, communication and content production are increasingly complex. Increased access to tools for creative generation and knowledge production will bring benefits to society. However, to realize this, one key component is transparency across the pipelines of AI design, content production and information distribution.[21] Transparency approaches can also support better control for individuals and others on how their data is used in AI models.

Frequently in my work, I am asked to provide advice to consumers on how to spot an AI-generated image—for example, to look for 'the distorted hands', or in the case of a deepfake, to see if it does not blink. I discourage this as these heuristics are the current Achilles heel or temporary failings of a process, not long-term durable or scalable guidance. Most audiovisual content we create and consume involves AI. In a world with wider access to tools that simplify the generation or edition of photos, videos, and audio, including photo and audio-realistic content, it is important for the public to be able to understand if and how a piece of media was created or altered using AI. We refer to this as 'content provenance'. Such labeling, watermarking or indications of provenance are not a punitive measure to single out AI content or content infused with AI, and should not be understood as a synonym of deception, misinformation or falsehood. The vast majority of synthetic media is

---

[18] Robert Chesney and Danielle Keats Citron, Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security, 107 California Law Review 1753, July 2018, *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3213954*

[19] Karen Hao, *China, a Pioneer in Regulating Algorithms, Turns Its Focus to Deepfakes,* Wall Street Journal, January 2023 *https://www.wsj.com/articles/china-a-pioneer-in-regulating-algorithms-turns-its-focus-to-deepfakes-11673149283*

[20] Sam Gregory, Journalism, December 2021, *ibid.*

[21] Sam Gregory, *Synthetic media forces us to understand how media gets made,* Nieman Lab, December 2022, *https://www.niemanlab.org/2022/12/synthetic-media-forces-us-to-understand-how-media-gets-made/*

used for personal productivity, creativity or communication without malice. Satirical media made using AI is also a critical and protected form of free speech.[22]

We have heard repeatedly from information consumers around the world that responsibility should not be placed primarily on end-users to determine if the content they are consuming is AI-generated, created by users with another digital technology or, as in most content, a mix of both.[23] To ensure disclosure—and more broadly, to promote transparency and accountability—all actors across the AI and media distribution pipeline need to be engaged. These include:

- Those researching and building foundation or frontier models;
- Those commercializing generative AI tools;
- Those creating synthetic media;
- Those publishing, disseminating or distributing synthetic media (such as media outlets and platforms); and
- Those consuming or using synthetic media in a personal capacity

There is now a significant trend in AI governance towards a pipeline approach and a focus on labeling and disclosure. In July 2023, seven leading AI companies agreed with the White House to a number of voluntary commitments to help move toward safe, secure and transparent development of AI technology, including committing to earning people's trust by disclosing when content is AI-generated.[24] In the European Union, companies who have signed on to the voluntary EU Code of Practice on Disinformation have agreed to a similar commitment, with the EU's Commissioner Věra Jourová calling on these companies to label AI-generated content.[25] The EU AI Act includes significant requirements for disclosing deepfakes and machine-generated content from foundation models.

Most provenance systems will require methods that explain both AI-based origins or production processes, but also document non-synthetic audio or visual content generated by users or other digital processes–like footage captured from 'old fashioned' mobile devices.[26] As the White House notes in its statement on the voluntary commitments, "companies making this commitment pledge to work with industry peers and standards-setting bodies as appropriate towards developing a technical framework to help users distinguish audio or visual content generated by users from audio or visual content generated by AI." It will be hard to address AI content in isolation from this broader question of media provenance. Implementing this approach to transparency will require standards that focus on how to design durable, machine-readable shared standards that provide useful signals to consumers, as well as other actors in the information pipeline (*e.g.,* content distributors and platforms).

The opportunity in transparency and disclosure

It is crucial for democracy that people are able to believe what they see and hear when it comes to critical government, business and personal communications, as well as documentation of events on the ground. It is also critical for realizing the creativity and innovation potential that generative AI holds, that consumers are informed about what they see and hear.

WITNESS has actively participated in the Partnership on AI's Responsible Practices for Synthetic Media Framework.[27] This Framework describes *direct* forms of

---

[22] Henry Ajder and Joshua Glick, *Just Joking! Deepfakes, satire, and the politics of synthetic media,* WITNESS and MIT, December 2012, *https://cocreationstudio.mit.edu/just-joking/*

[23] WITNESS, *Synthetic Media, Generative AI And Deepfakes Witness' Recommendations For Action,* 2023, *https://www.gen-ai.witness.org/wp-content/uploads/2023/06/Guiding-Principles-and-Recs-WITNESS.pdf*

[24] The White House, *FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI,* July 2023 *https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/* and The White House, *Ensuring Safe, Secure, and Trustworthy AI, https://www.whitehouse.gov/wp-content/uploads/2023/07/Ensuring-Safe-Secure-and-Trustworthy-AI.pdf*

[25] Foo Yun Chee, *AI-generated content should be labelled, EU Commissioner Jourova says,* Reuters, June 2023 *https://www.reuters.com/technology/ai-generated-content-should-be-labelled-eu-commissioner-jourova-says-2023-06-05/*

[26] Sam Gregory, *To battle deepfakes, our technologies must track their transformations,* The Hill, June 2022, *https://thehill.com/opinion/technology/3513054-to-battle-deepfakes-our-technologies-must-lead-us-to-the-truth/*

[27] Partnership on AI, *Responsible Practices for Synthetic Media Framework, https://syntheticmedia.partnershiponai.org/* See also, Jacobo Castellanos, *Building Human Rights Ori-*

disclosure as those methods that are 'visible to the eye', such as labels marking the content, or adding context disclaimers. *Indirect* forms of disclosure are not perceptible to the human eye and include embedded metadata or other information that is machine readable or presentable, such as cryptographic provenance or embedding durable elements into either or both the training data and the content captured or generated.[28] Importantly, the Framework also offers a useful breakdown of how responsibility for supporting this disclosure should be considered at different stages across the AI pipeline.

There is significant and unhelpful confusion around terms used to show use of AI in content.[29] 'Watermarking' is used as a catch-all term that includes:

a. Visible watermarks, signals or labels (*e.g.,* a 'Made with AI' description on an image);
b. Invisible watermarks and technical signals that are imperceptible to the human eye and can be embedded at pixel level, or as early as the training stage of AI processes; and
c. Cryptographically-signed metadata that shows the production process of content over time, like the C2PA standards.

Visible signals or labels can be useful in specific scenarios such as AI-based imagery or production within election advertising, as proposed in the REAL Political Ads Act. However, visible watermarks are often easily cropped, scaled out, masked or removed, and specialized tools can remove them without leaving a trace. Visible watermarks are hence inadequate for reflecting the 'recipe' for the use of AI in an image or video, and in a more complex media environment fail to reflect how AI is used in a meaningful way for consumers.

Technical interventions and signals at the dataset level can indicate provenance as well as embed 'Do Not Train' restrictions that could give consumers more say in who is allowed to build AI models using people's data and content. However, many datasets are already in use and do not include these marks. Additionally, small companies, independent developers, and open-source libraries and tools may not have the capacity and ability to develop reliable and sustainable invisible watermarks. Without accessible and standardized standards, there is a risk of excluding a significant part of the AI innovation ecosystem from the conversation. This, in turn, could lead to a handful of AI companies' dominance becoming further entrenched. Dataset-level watermarks also require their application across broad datasets, which brings questions around ownership and responsibility regarding the content and the repurposing of that content for training purposes. Since, in most cases, the original content creators are not involved in the decision to add their content to a training dataset, they are unlikely to be involved in the decision to watermark their content as well.

Cryptographic signature and provenance-based standards such as the C2PA are built to make it very hard to tamper with the cryptographic signature without leaving evidence of the attempt, and to enable the reconnection of a piece of content to a set of metadata if that is removed. These methods can allow people to understand the lifecycle of a piece of content, from its creation or capture to its production and distribution. In some cases, they are integrated with capture devices such as cameras, utilizing a process known as 'authenticated capture'. Microsoft has been working on implementing provenance data on AI content using C2PA specs,[30] and Adobe has started to provide it via its Content Credentials approach.[31] While I do not speak for the C2PA, WITNESS is a member of the C2PA, has participated in the Technical Working Group and acted as a co-chair of the C2PA Technical Working Group Threats and Harms Taskforce. In this context WITNESS has advocated for

---

ented Guidelines for Synthetic Media, WITNESS, February 2023, *https://blog.witness.org/2023/02/building-human-rights-oriented-guidelines-for-synthetic-media/*

[28] For synthetic content, the most recent example is SynthID, released by Google on August 29, 2023. *https://www.deepmind.com/blog/identifying-ai-generated-images-with-synthid*

[29] Claire Leibowicz, *Why watermarking AI-generated content won't guarantee trust online,* MIT Tech Review, August 2023, *https://www.technologyreview.com/2023/08/09/1077516/watermarking-ai-trust-online/*

[30] Kyle Wiggers, *Microsoft pledges to watermark AI-generated images and videos,* Techcrunch, May 2023 *https://techcrunch.com/2023/05/23/microsoft-pledges-to-watermark-ai-generated-images-and-videos*

[31] Adobe Content Credentials, *https://helpx.adobe.com/creative-cloud/help/content-credentials.html*

globally-driven human rights perspectives and practical experiences to be reflected in the technical standard.[32]

An approach like the C2PA can also allow creators to choose whether their content may be used for training AI models or other data purposes. Agency over their data is a critically needed response to concerns from creators and others about the incorporation of their content, personal images and other data into AI models without their consent.

In reality, any effective shared standard, regulation, or technological solution to provenance, disclosure and transparency is likely to require a combination of cryptographically-signed provenance metadata that reflects how both AI, non-AI and mixed media are created and edited over time, as well as visible watermarking and/or technical signals for synthetic content that confirm the use of AI specifically.

How to provide rights-respecting disclosure

To safeguard Constitutional and human rights, approaches to provenance and disclosure need to meet at least three core criteria. They need to:

- Protect privacy:
- Be accessible with modular opt-in or out depending on the type of media and metadata, and;
- Avoid configurations that can be easily weaponized by authoritarian governments.

People using generative AI tools to create audiovisual content should not be required to forfeit their right to privacy to adopt these emerging technologies. Personally-identifiable information should not be a prerequisite for identifying either AI-synthesized content or content created using other digital processes. The 'how' of AI-based production elements is key to public understanding; this should not require a correlation to the identity of 'who' made the content or instructed the tool.

Since 2019, WITNESS has been raising concerns about the potential harms that could arise from the inclusion of personal data in solutions that track the provenance of media.[33] The U.S. government has the opportunity to ensure that provenance requirements and standards are developed in-line with global human rights standards, protect civil rights and First Amendment rights, and do not include the automated collection of personal data. While a requirement to include disclosure indicating content was AI-generated could be a legal requirement in certain cases, this obligation should not extend to using tools for provenance on content created outside of AI-based tools, which should always be opt-in.

Building trust in content must allow for anonymity and redaction. Immutability and inability to edit do not reflect the realities of people, or how and why media is made–nor that certain redaction may be needed in sensitive content.[34] Flexibility to show how media evolves—and to conduct redaction—is a functional requirement for disclosure particularly as it relates to edited and produced content. Lessons from platform policies around 'real names' tell us that many people—for example, survivors of domestic violence—have anonymity and redaction needs that we should learn from.[35] While specifications like the C2PA focus on protecting privacy and don't mandate identity disclosures, this privacy requirement needs to be protected during widespread adoption. We should be wary of how these authenticity infrastructures could be used by governments to capture personally identifiable information to augment surveillance and stifle freedom of expression, or facilitate abuse and misuse by other individuals.

We must always view these credentials through the lens of who has access and can choose to use them in diverse global and security contexts, and ensure they are accessible and intelligible across a range of technical expertise.[36] Provenance data for both AI and user-generated content provides signals—i.e. additional information about a piece of content—but does not prove truth. An 'implied truth' effect simply

---

[32] The Coalition for Content Provenance and Authenticity, *C2PA Harms Modelling, https://c2pa.org/specifications/specifications/1.0/security/Harms_Modelling.html;* Jacobo Castellanos, WITNESS, 2021, *Ibid.*

[33] Gabriela Ivens and Sam Gregory, *Ibid;* Sam Gregory, *Tracing trust: Why we must build authenticity infrastructure that works for all,* WITNESS, 2020, *https://blog.witness.org/2020/05/authenticity-infrastructure/*

[34] Raquel Vazquez Llorente, *Trusting Video in the Age of Generative AI,* Commonplace, June 2023, *https://commonplace.knowledgefutures.org/pub/9q6dd6lg/release/2*

[35] Jillian York and Dia Kayyali, *Facebook's 'Real Name' Policy Can Cause Real-World Harm for the LGBTQ Community,* EFF, 2014, *https://www.eff.org/deeplinks/2014/09/facebooks-real-name-policy-can-cause-real-world-harm-lgbtq-community*

[36] Sam Gregory, *Ticks Or It Didn't Happen,* WITNESS, December 2019, *https://lab.witness.org/ticks-or-it-didnt-happen/*

derived from the use of a particular technology is not helpful, nor is an 'implied falsehood' effect from the choice or inability to use them.[37] Otherwise we risk discrediting a citizen journalist for not using tools like these to assert the authenticity of their real-life media because of security or access concerns, while we buttress the content of a foreign state-sponsored television channel that does use it. Their journalism can be foundationally unreliable even if their media is well-documented from a provenance point of view.

Any credential on content must be an aid to help make informed decisions, not a simplistic truth signal. They work best as a signal in complement to other processes of digital and media literacy that consumers choose to use, to help them triage questions they may have, and that are available to other parties engaging with the content, including potentially platforms.

The role of detection alongside provenance

'Seeing' both invisible watermarks and provenance metadata that are imperceptible to the eye will require consumer-facing tools. However, the average citizen shouldn't be required to keep up with watermarking advances and detection tools, and cannot be expected to deploy multiple tools to ascertain if a particular commercial brand, watermarking approach, or mode of synthesis has been used.

Detection tools are also necessary for content believed to be AI-generated that does not have provenance information or that has been manipulated with counter-forensics approaches. There is justifiable skepticism about whether after-the-fact detection tools are useful for consumer transparency and consumer usage to identify generative AI and deepfake outputs.[38] Detection of audiovisual generative AI and deepfakes outputs is flawed. Existing detection models frequently require expert input to assess the results and often they are not generalisable across multiple synthesis technologies and techniques or require personalization to a particular person to be protected from fraudulent voices or imagery. As such, detection tools can lead to unintentional confusion and exclusion. We have seen how use by the general public of detection tools has contributed to increased doubt around real footage and enabled the use of the liar's dividend and plausible deniability around real content, rather than contributing to clarity.[39]

However, from WITNESS's experience they are a critical element—alongside the incorporation of provenance data and media literacy—when it comes to real-world scenarios where journalists, civil society and governments are attempting to discern how content has been created and manipulated. As we have seen in our work supporting forensic analysis of high profile global cases, there is a gap between on one side the needs of journalists, civil society leaders and election officials, and on the other side the availability of detection skills, resources and tools that are timely, effective and grounded in local contexts. These issues highlight the 'detection equity' gap that exists—the tools to detect AI-generated media are not available to the people who need them the most. Further research into improving detection capabilities remains critical as well as ensuring those who access tools also have the knowledge and skills to use them.

Conclusion

Significant evolutions in volume, ease of access, personalization and malicious usage of generative AI reflect both the potential for creativity but also the heightened harms from audiovisual generative AI and deepfakes–including the plausible deniability that these tools enable, undermining consumers' trust in the information ecosystem. I have highlighted in this statement the need to focus on existing harms as identified by those on the frontlines of deepfakes and synthetic media, and to center the voices of those affected by an AI-powered information landscape. I encourage this Subcommittee and legislators to go beyond a risks-based approach, and push for a rights-based framework in order to prevent and mitigate accidental harms, misuse harms, supply chain harms and structural harms. In this regard, approaches to transparency in audiovisual content production that incorporate strong

---

[37] Sam Gregory, Journalism, December 2021, *ibid.*

[38] Riccardo Corvi, Davide Cozzolino, Giada Zingarini, Giovanni Poggi, Koki Nagano, Luisa Verdoliva, *2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), On The Detection of Synthetic Images Generated by Diffusion Models, https://arxiv.org/abs/2211.00680;* Luisa Verdoliva, *Media Forensics and Deepfakes: An Overview,* 2020, *https://arxiv.org/abs/2001.06564*

[39] Sam Gregory, *Pre-Empting a Crisis: Deepfake Detection Skills + Global Access to Media Forensics Tools,* WITNESS, *https://blog.witness.org/2021/07/deepfake-detection-skills-tools-access/;* Nilesh Christopher, *ibid;* Sam Gregory, *The World Needs Deepfake Experts to Stem This Chaos,* WIRED, June 2021, *https://www.wired.com/story/opinion-the-world-needs-deepfake-experts-to-stem-this-chaos/*

privacy measures can protect personal information, safeguard democracy around the world, and promote creative production.

Senator HICKENLOOPER. Thank you, Mr. Gregory.
Mr. Stair—Strayer, I am sorry.

STATEMENT OF ROB STRAYER,
EXECUTIVE VICE PRESIDENT OF POLICY,
INFORMATION TECHNOLOGY INDUSTRY COUNCIL (ITI)

Mr. STRAYER. Thank you. Chair Hickenlooper, Ranking Member Blackburn, and Members of the Committee; thank you for the opportunity to testify today.

My name is Rob Strayer, and I lead the Global Policy Team, at the Information Technology Industry Council, or ITI. ITI represents companies from all corners of the technology sector, and from across the AI ecosystem, including those involved in both developing AI models, and deploying cutting-edge AI applications.

ITI was pleased to provide a very detailed response to the Subcommittee's inquiry earlier this year, about how our companies are operationalizing the NIST AI Risk Management Framework, as a means of building public trust.

We are encouraged by the bipartisan efforts in Congress to address the challenges and opportunities from AI. In my remaining time, I will address the competitive, global context for AI, and then turn to transparency and accountability.

The development and adoption of AI technology will be transformational across all sectors of the economy, estimates the total global economic benefits of AI in the years ahead range from $14 trillion to $25 trillion. That is absolutely massive.

And just one example of the transformational nature of AI, the cybersecurity industry is able to deploy machine learning to detect and stop the most sophisticated attacks, using zero-day exploits. AI can defeat these pernicious attacks using insights about activity rather than having to rely only on known malware signatures.

Adversaries will certainly be using AI to improve their attacks. And we need to be able to leverage AI to protect our critical infrastructure and IT systems; AI also will play an essential role in future national security applications for the military and for the intelligence community.

The good news is that today, the United States is leading AI development, deployment, and innovation globally. Nonetheless, foreign competitors are working hard on AI breakthroughs and to deploy AI and new use cases in their markets. And with access to open source models, and decreasing model training compute costs, AI developers and deployers will be able to move anywhere in the world with interconnections to avoid stifling regulations.

Therefore, U.S. policymaking involving AI needs to be understood in a global context to consider how new policies affecting AI will help the United States maintain its technological leadership rather than cede it to competitors, including authoritarian states.

So how does the United States create a pro-innovation AI policy framework that manages risk? U.S. AI policies should have two primary components; one, promoting innovation and investment, and two, building public trust and accountability. My written testi-

mony covers the importance of investment, and so I will focus on public trust and accountability here.

Transparency is a key means by which to establish public trust. Consumer trust will increase adoption of AI and expand the AI ecosystem in the United States. ITI companies are working to ensure that users understand when they are interacting with an AI system, and generally how the system works. ITI companies also are producing AI model cards, so that consumers have access to the information about features and limitations of AI models in clear, plain language.

So what is the government's role? To avoid regulations being overly broad, risk should be identified in the context of a specific AI use case. With risk identified it is then imperative that the government review the existing regulatory landscape. Legal regimes such fraud, criminal law, as well as statutes like the Civil Rights Act, can address AI-related risk.

It is critical to understand how these legal regimes function, and where they may not be fit for purpose to address AI risk before creating new legislation or regulatory frameworks.

Finally, before adopting new AI regulatory requirements, policymakers should understand the status of international consensus-based standards, and the ability of those standards to meet regulatory requirements. Without specific standards for risk management processes, such as the measurement and evaluation of the risk in models, it will not be possible to implement regulations effectively, or harmonize rules globally.

To wrap up, Congress and private sector stakeholders can work together to ensure the United States builds on its competitive lead in AI, as AI transforms all sectors of the economy, and generates trillions of dollars in economic growth. This will benefit U.S. companies and citizens for decades into the future.

Thank you. And I look forward to your questions.

[The prepared statement of Mr. Strayer follows:]

PREPARED STATEMENT OF ROB STRAYER, EXECUTIVE VICE PRESIDENT OF POLICY, INFORMATION TECHNOLOGY INDUSTRY COUNCIL (ITI)

Chairman Hickenlooper, Ranking Member Blackburn, and Distinguished Members of the Committee and Subcommittee, thank you for the opportunity to testify today.

My name is Rob Strayer, and I'm the Executive Vice President of Policy at the Information Technology Industry Council (ITI). I lead ITI's global policy team, driving ITI's strategy and advocacy efforts to shape technology policy around the globe and enable secure innovation, competition, and economic growth, while supporting governments' efforts to achieve their public policy objectives. ITI is the premier advocate and thought leader in the United States and around the world for the information and communications technology (ICT) industry. We represent leading companies from across the ICT sector, including hardware, software, digital services, semiconductor, network equipment, cybersecurity, Internet companies, and other organizations using data and technology to evolve their businesses. Our members stand at the forefront in developing and deploying consumer-facing, business-to-business, and government-focused AI solutions.

We are encouraged by the bipartisan efforts in Congress to address the challenges and opportunities from AI. This subcommittee's jurisdiction over issues ranging from data privacy and consumer protection to standards gives you an important role to play in AI policy discussions. To that end, ITI was pleased to provide information to Chairman Hickenlooper and Ranking Member Blackburn's inquiry earlier this year about how ITI members are operationalizing NIST's AI Risk Management

Framework (AI RMF) to build and foster public trust.[1] Congress and the Administration should work together to ensure any legislation or regulatory proposals encourage future innovation and investment in the United States, protect consumers and businesses, mitigate foreseeable risks, and do not complicate or duplicate existing standards, laws, and sector-specific regulatory regimes. ITI looks forward to being a partner in those efforts.

I. Transformational Impact of AI

The development and adoption of AI technologies will be transformational across a variety of critical sectors, including health care, telecommunications, aerospace, manufacturing, transportation, and other sectors under the Committee's jurisdiction. It will help companies be more effective and efficient, particularly at addressing business operations challenges, research and development, and software engineering. In fact, an Accenture survey of 1,500 executives across all sectors found that 84 percent believed AI is critical to meeting their growth objectives and 73 percent said they risk going out of business if they cannot scale AI.[2]

As a testament to AI's revolutionary impact, credible estimates of the total global economic benefits of AI in the years ahead, which now includes the impact of generative AI, range from $14 trillion to $25 trillion.[3]

Today, the United States is leading AI development, deployment, and innovation. The United States employs the best and the brightest AI researchers and experts working to advance American leadership in AI innovation. Other nations have recognized the United States as the center for AI excellence and are working harder than ever to develop the next major technological developments in AI and to deploy AI in new use cases in their countries.

Policymaking and regulation involving AI needs to be understood in the global context of technology competition. *The United States has the potential to build on its lead as AI transforms all sectors of the economy, generates trillions of dollars in economic growth, and benefits U.S. companies and citizens for decades into the future. Overly broad and prescriptive regulation, however, could undermine that leadership position and cede it to U.S. competitors, including authoritarian nations.*

AI will play an essential role in future national security applications for the military and intelligence communities and in the cybersecurity defense of critical infrastructure. It is not an exaggeration to say that U.S. national security depends on continued U.S. technological leadership in AI. It is more important than ever that the United States considers how any new policy affecting AI will help it maintain its technological leadership in AI.

Below are some of the use cases that AI will empower:

- *Cybersecurity*
  - *Threat Mitigation:* AI and machine learning can be leveraged to improve cybersecurity. Indeed, defensive cybersecurity technology must embrace machine learning and AI as part of the ongoing battle between attackers and defenders. The threat landscape constantly evolves, with cyberattacks that are complex, automated and constantly changing. Attackers continually improve their sophisticated and highly automated methods, moving throughout networks to evade detection. The cybersecurity industry is innovating in response: making breakthroughs in machine learning and AI to detect and block the most sophisticated malware, network intrusions, phishing attempts, and many more threats.[4] AI is the best tool defenders have to identify and prevent zero-day attacks and malware-free attacks because AI can defeat novel threats based on behavior cues rather than known signatures. Leveraging these technologies is essential to meeting constantly evolving threats.

---

[1] ITI's June 2023 response to Chairman Hickenlooper and Ranking Member Blackburn's April 2023 letter (June 1, 2023) available at: *https://www.itic.org/documents/artificial-intelligence/ITIJune2023ResponsetoSens.HickenlooperandBlackburnAIRMFLetter.pdf*

[2] *See* Accenture AI investment study (November 14, 2019), available at *https://www.accenture.com/us-en/insights/artificial-intelligence/ai-investments*

[3] *See* McKinsey and Company *The Economic Potential of Generative AI: The Next Productivity Frontier* (June 2023), available at *https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/the%20economic%20potential%20of%20generative%20ai%20the%20next%20productivity%20frontier/the-economic-potential-of-generative-ai-the-next-productivity-frontier-vf.pdf?shouldIndex=false*

[4] Testimony of Rob Strayer, Hearing on Securing the Future: Harnessing the Potential of Emerging Technologies while Mitigating Security Risks, Before the U.S. House Homeland Security Committee (June 22, 2022) available at *https://www.itic.org/documents/cybersecurity/20220622ITIHouseHomelandCmteTestimonyonEmergingTechandCyber.pdf*

- *Manufacturing*
  - *Predictive Maintenance:* AI can analyze real-time sensor data from manufacturing equipment to predict maintenance needs accurately. By identifying potential equipment failures in advance, manufacturers can schedule maintenance proactively, minimizing unplanned downtime and reducing costs.
  - *Supply Chain Management:* AI can optimize supply chain operations by analyzing data from multiple sources, including demand forecasts, inventory levels, and logistical constraints. AI algorithms can optimize inventory management, improve demand forecasting accuracy, and enable efficient routing and scheduling of shipments.
- *Health Care*
  - *Medical Imaging Analysis:* AI can analyze medical images such as X-rays, CT scans, and MRIs, helping doctors detect and diagnose diseases more accurately and efficiently. AI can assist in identifying anomalies, tumors, or other abnormalities, leading to earlier detection and treatment.
  - *Drug Discovery and Development:* AI accelerates the drug discovery process by analyzing massive datasets and identifying potential drug candidates. AI algorithms can predict the efficacy of drugs, design molecules, and optimize clinical trials, reducing the time and cost of bringing new drugs to market.
- *Telecommunications*
  - *Network Planning and Deployment:* AI can help analyze data to assist in planning the deployment of telecommunication networks. AI can help determine optimal tower locations, estimate coverage areas, and predict network capacity requirements, enabling providers to make informed decisions during network expansion.
  - *Network Security:* AI can monitor network traffic, detect anomalies, and identify potential cybersecurity threats. AI algorithms can analyze patterns, identify malicious activities, and take immediate action to protect the network and customer data from cyberattacks.

The United States needs to develop a pro-innovation policy framework that appropriately manages risk while maintaining U.S. technological leadership.

ITI's AI policy framework has four key pillars: (1) fostering innovation and investment, (2) facilitating public trust in and understanding of the technology, (3) ensuring security and privacy, and 4) maintaining global engagement.[5] My testimony today primarily focuses on the first two, although a comprehensive framework should seek to address all the above policy pillars.

A. A pro-innovation policy framework should support innovation and investment.

While much of the conversation of late has focused on ways in which to foster accountability, *there needs to be at least equal attention given to fostering innovation and investment.* Continued investment in AI research and development, by both the government and private sector, is essential for the United States to maintain its leadership position. ITI applauds the funding and authorizations in the CHIPS and Science Act for Federal efforts to enhance U.S. technological leadership in AI and other emerging technologies.

Regulatory policies that encumber the ability of researchers and developers in the United States will drive investments and research activities into other countries. Through open-source models and platforms, access to AI capabilities will be placed increasingly in the hands of a growing number of innovators of all sizes. This combined with decreasing costs for AI compute training resources, which are estimated to decrease 70 percent annually, will allow innovators to migrate away from jurisdictions with stifling regulations.

Most funding for AI research and development will come from the private sector. Smart technology investment-related tax policies and market incentives can encourage greater investments by the private sector that will produce AI innovations. ITI has detailed its views on these tech policies elsewhere.[6]

Government sponsored research in AI also has a role. Government investments in foundational science and AI-specific program research are important to fill gaps. Research by academia and the private sector into privacy enhancing technologies

---

[5] See ITI Global AI Policy Principles, available at *https://www.itic.org/documents/artificial-intelligence/ITI_GlobalAIPrinciples_032321_v3.pdf*

[6] See, *e.g., https://www.itic.org/news-events/techwonk-blog/congress-must-act-to-support-us-research-and-development*

(PETs) and in measurement science to test, evaluate, validate, and verify (TEVV) model performance are critical to effectively implementing a risk management approach. *Innovations in measurement tools for AI will make risk management more concrete and objective and improve accountability and transparency.*

The government also has a role in incentivizing professional and technical apprenticeships, education and training programs in STEM fields, and promoting access to external and online reskilling programs. AI is not just a function of STEM or advanced technical training; the best way to ensure access to an AI workforce is to invest broadly across all relevant disciplines and teach flexible skills and problem solving from early childhood education.

B. A pro-innovation policy framework should be risk-based, evaluate the existing regulatory landscape, and clearly delineate risk areas that are not adequately addressed.

In seeking to support innovation, it is important that we understand the risks that we are seeking to address with a regulatory framework. AI will continue to evolve, and we need to address risks as they develop, while not suppressing the advancement of AI.

Risks need to be identified and mitigated in the context of the specific AI use. This will help policymakers determine use cases or applications that are of particular concern, avoiding overly prescriptive approaches that may serve to stifle innovation. Beyond that, context is key. Not all AI applications negatively impact humans, and thus, they cannot inflict harm that would warrant regulation.

With those risks identified, the next step is to consider the role for existing statutory and regulatory authorities to address discrete risk. We don't want layers of regulation that conflict with one another, create undue burdens on innovators, and slow advancement.

Therefore, it is imperative that the government review the existing regulatory landscape to assess where there might be gaps. *There are existing laws and regulatory frameworks that can address AI-related risks, so it is critical to understand how those laws apply, and where they may not be fit-for-purpose, prior to creating new legislation or regulatory frameworks pertaining to AI.*

As an initial step, policymakers should evaluate how NIST's AI RMF is being adopted and how it can be used to manage risk. The AI RMF provides companies with a comprehensive way to think about risk management practices, which is fundamental to fostering long-term public trust. It captures many of the outcomes and best practices that companies are already undertaking, such as framing and prioritizing risks and addressing AI trustworthiness characteristics (*e.g.,* reliability, safety, explainability, privacy, fairness, accountability, and transparency). ITI and its member companies were active in the development of this Framework and are actively adopting it. We appreciate that NIST has also launched the AI RMF Playbook as a complement to the AI RMF. Indeed, this tool is instrumental to ensuring that the Framework is actionable and implementable, particularly for organizations that may be less familiar with the scope of guidelines and best practices that are available to them. In recent comments to the Office of Science and Technology Policy, we encouraged the Administration to explore how the AI RMF might be integrated into Federal contracts and encouraged the government to leverage the AI RMF in crafting forthcoming guidance. [7]

Conducting a robust gap analysis of existing legal authorities relevant to AI's potential harms is critical because there are many laws and regulations that can address the diversity of impacts implicated by the technology. Some of these relevant bodies of law and regulation, coupled with relevant potential AI-related harms, include: intellectual property law, especially the Copyright Act of 1976, to address issues related to the use of copyrighted material in training data and questions regarding the IP rights in AI generated content; the Federal Trade Commission Act to address unfair, deceptive or abusive practices related to AI-enabled misrepresentations or harmful content; product liability common law to address potential safety issues related to products containing AI technology that may cause physical injury; First Amendment jurisprudence and Section 230 of the Communications Decency Act to address issues related to AI-generated content and freedom of expression interests; Title VII of the Civil Rights Act of 1964 and related laws to address issues related to bias, discrimination, or other civil rights harms; and relevant Federal sector-specific privacy provisions, such as in the Health Insurance Portability and Ac-

[7] *See* ITI's July 2023 response to OSTP RFI, re: ITI Response to Office of Science and Technology Policy Request for Information on National Priorities for Artificial Intelligence, available at *https://www.itic.org/documents/artificial-intelligence/ITIResponsetoOSTPRFIonNationalAIPrioritiesFINAL%5B25%5D.pdf*

countability Act, to address potential privacy harms related to AI that include the accuracy of data.

In our view, it makes sense to proceed with creating new legislation only if there is a specific harm or risk where existing legal frameworks are either determined to be insufficient or do not exist.

Regarding privacy protections, as noted above, privacy laws do exist that can address some AI-related privacy harms. Yet, there is also an undeniable regulatory gap given the absence of Federal privacy legislation. ITI testified before the House Energy and Commerce Committee in 2022 in favor of preemptive Federal comprehensive privacy legislation, which we consider critical to protecting consumers from data related harms and a necessary complement to any potential AI legislation or regulation.[8] However, ITI urges the Committee not to conflate potential AI legislative provisions with comprehensive privacy legislation. For example, in ITI's testimony on the American Data Privacy and Protection Act, we expressed concerns that the bill conflated the two issues by prematurely including prescriptive requirements to conduct algorithmic design evaluations and impact assessments, and that the scope of those requirements, which would have potentially covered all algorithms, were overbroad and would have swept in a vast array of technologies well beyond AI.

C. A pro-innovation policy framework should aim to foster public trust in the technology.

The guiding goal of an AI policy or regulatory framework should be fostering public trust in AI technology. Fostering trust in AI systems[9] requires AI model developers, deployers, and policymakers to work together. If we are successful in achieving that trust, adoption of AI by consumers and businesses will increase. AI adoption will benefit the users of new services, and it will encourage further development and experimentation in AI and other emerging technology fields, such as quantum and high-performance computing. Commercial successes will provide resources to companies that they can invest in AI and other innovations. In that way, *trust is an essential element of a beneficial research and development cycle for technology and the expansion of the AI ecosystem.*

Transparency is a key means by which to achieve that trust. To support those efforts, ITI developed *AI Transparency Policy Principles.*[10] Indeed, ITI members are actively taking steps to build and deploy safe and transparent AI technologies for products and systems. Transparency is paramount for our member companies, particularly when it comes to fostering trust in AI technology. They have placed a premium on these activities. While transparency can take different forms, our companies are working to ensure that users understand when they are interacting with an AI system and broadly how that system works.

In general, transparency can be understood as being clear about how an AI system is built, operates, and functions. When appropriately configured, transparency mechanisms can help to comprehend outputs of an AI system and foster accountability. Transparency is an overarching concept, with both explainability and disclosure falling under this umbrella. In contemplating policy approaches to transparency, we highlight several key considerations that legislators should consider.

First, like with policy approaches to AI generally, *transparency requirements should be risk-based.* It is important to consider the diversity of possible AI use cases and applications, given that the demand for transparency requirements from various users may vary significantly based on the AI application or intended use. Many use cases present little to no risk to the user, and so imposing transparency requirements in such situations will likely add little value to the user and hinder innovation by adding onerous, disproportionate requirements.

Second, in thinking about transparency, *it is important to consider the objective and intended audience.* The target audience at which transparency requirements are directed, including their level of expertise, plays a key role. For example, trans-

---

[8] Testimony of John Miller, Hearing on Protecting America's Consumers: Bipartisan Legislation to Strengthen Data Privacy and Security, Before the U.S. Energy and Commerce Committee (June 14, 2022) available at *https://docs.house.gov/meetings/IF/IF17/20220614/114880HHRG-117-IF17-Wstate-MillerJ-20220614.pdf*

[9] We define an *AI system* as a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy. This is based on the OECD definition of AI.

[10] *https://www.itic.org/documents/artificial-intelligence/ITIsPolicyPrinciplesforEnablingTransparencyofAISystems2022.pdf.* Transparency is not the only means by which to foster public trust, and so any approach should consider the role that transparency can play, as well as other tools.

parency could be useful to several different audiences (*e.g.,* regulators, consumers, developers, etc.), which will in turn influence the development of potential provisions. Understanding the intended audience will also inform the type of information presented, the way it is presented, and the amount of information presented.

Third, *disclosure can play an important role in facilitating transparency.* AI systems should be disclosed when they are playing a significant role in decision-making or interacting directly with users. In the context of disclosure, language should be plain and clear so that it is understandable to a wide audience. Disclosures should generally include high-level information, such as a topline explanation of how an AI system works, capabilities, and known limitations on performance. Additionally, disclosure should be the responsibility of the deployer to ensure that disclosure and other consumer-facing obligations are met. That said, the developer of the AI system should ensure that terms of sale do not prohibit disclosure. Relatedly, we are supportive of information-sharing in the value chain to facilitate cooperation between developers and deployers.

Finally, *Congress should avoid an overly prescriptive approach to transparency and maintain appropriate IP protections.* It is important that transparency requirements allow for flexibility because it may not be appropriate or useful to provide the same type of details in every context or for every target audience.[11] Organizations should have the ability to tailor such information, depending on context, use of, and level or risk associated with the system. Also important is that transparency requirements do not require the disclosure of source code and sensitive IP, or otherwise reveal sensitive individual data. Any requirements around transparency should avoid requiring companies to divulge sensitive IP or source code. Disclosure of source code could seriously put at risk trade secrets, undermine IP rights, and contravene widely accepted best practices for digital trade. It could also pose risks to safety and security and allow malicious actors to manipulate an AI system.

D. A pro-innovation policy framework should consider the views and input of all stakeholders.

Both developers and deployers of AI systems should be consulted in the development of policy frameworks, as well as civil society, academia, and companies operating across different sectors. Small, medium, and large companies in all sectors will be using AI to be more efficient and offer better quality services and products, so it is imperative to cast a wide net to obtain a diverse set of perspectives.

Additionally, a pro-innovation policy framework should seek to appropriately delineate the roles and responsibilities of different stakeholders in the AI value chain. *Stakeholders throughout the value chain, including small businesses across a variety of sectors, play a role in the development and deployment of AI in a responsible manner.* As such, responsibilities should reflect the important distinction between developers and deployers and be allocated among actors based on their role and function in the AI value chain.

E. A pro-innovation policy framework should rely upon globally-recognized international standards.

Before adopting new AI regulatory requirements, policymakers should understand the status of international consensus-based standards and the ability of those standards to meet regulatory requirements. AI standards are essential to increase interoperability, harmonization, and trust in Al systems. They can inform AI regulation, practical implementation, governance, and technical requirements. *Without specific standards for risk management processes, such as the measurement and evaluation of models, it will not be possible to implement regulations effectively.* Moreover, regulations that are not aligned with international standards will undermine the leadership of companies doing business in the United States that seek to scale into other jurisdictions that adopt international standards.

The International Standards Organization and International Electrotechnical Commission have formed a Joint Technical Committee for IT standards, and it has established a subcommittee on Artificial Intelligence (SC 42). That subcommittee has published several AI standards and is working on an Artificial Intelligence Management System (AIMS) standard that will cover processes for the development or use of Al, such as bias, fairness, inclusiveness, safety, security, privacy, accountability, explicability, and transparency.[12] This management system standard will help advance innovation and technology development through structured governance and appropriate risk management. SC 42 currently also has other standards under

[11] Using AI to limit fraud, spam, illegal, or malicious information are some examples of where including technical details or too prescriptive of a disclosure may be inappropriate.

[12] See ISO/IEC 42001 Information technology—Artificial intelligence—Management system.

development, focused on terminology, reference architecture, governance of AI, and trustworthiness. We encourage Congress to consider how standards can address risk management requirements and ensure international harmonization.

III. Conclusion

To lead in AI, the United States needs a pro-innovation policy framework that prioritizes innovation and investment as well as building public trust in AI. Public trust will increase AI adoption by businesses and consumers, and accelerate the flourishing of the AI ecosystem of developers and deployers.

Building public trust through risk management will be critical, and the private sector is already leading in this area, such as through adoption of the NIST AI RMF. The government's role should be limited to addressing critical risks in specific use cases. Where those risks are identified, Congress should evaluate the existing legal and regulatory landscape, and clearly delineate risk areas that are not adequately addressed, before enacting new legislation to address any gaps. Future requirements should be aligned with international consensus standards wherever possible to ensure that risk management is effective and to harmonize the global marketplace for technology.

Thank you, and I look forward to your questions.

Senator HICKENLOOPER. Great. Thank each of you for being here, for your opening statements, for all the work you have done on these issues already.

This first question I have might be considered obvious. AI systems learn about the world by processing written, visual, and audio files, works created by people, by humans.

Dr. Krishnan, just to kind of lay out, because a lot of people don't—or come into this issue fairly freshly; what rights already exist for consumers to decide how AI systems access their creations, and what additional rights do they have?

Dr. KRISHNAN. Can you hear me now? Okay. Thank you. Thank you for the question. So perhaps a place to start is to think about the AI pipeline first, and then I will focus in particular on the creators of content, and the need to appropriately balance what a typical creator's interests are, and how that might be protected.

So we think about the AI pipeline, it includes training data, models, and applications. When you have data that models use, that involves creative artifacts, be it music, images, video, that is copyrighted, a group of creators may actually want to seek advertising off this content that they have created, and therefore may pull such information on the web, with the full expectation that consumers interested in this content may sample it, and then they may be able to earn income from it.

The concern here is that if this content is then scraped and used in AI models, which then produce output in the style of the same kind of content that the creators have created, that could potentially take away revenue from the individual creators who have created this content. So this issue on the one hand of protecting creators, who would like to have income be generated from the creative acts that they are engaged in, on the one hand.

The other is the capacity of these models to use these types of data for the purposes of creating the capabilities that we have witnessed. So one potential path forward, on the one hand, you want to have to copyright, and the benefits that accrue from licensing that come from that, is perhaps to use technology, as work from the University of Chicago, that allows for individuals to upload their creative content, and the technology makes modifications for that content which is not visible to the human eye.

So the human sees it as an image, just like the artist intended it to, but it is not trainable by an AI model, so that the AI model can't produce it in the likeness of the artist. And if the model developer wants to obtain access to that content they can licenses it, and that might be, potentially, a way of, on the one hand, providing notice and disclosure, which currently doesn't exist through those people who have created this content, whose content got scraped, while at the same time, meeting the needs both of the model developer, and of the artist.

Senator HICKENLOOPER. Right. Got it.

Mr. Gregory, what would you add to that, as the next step? That is a broad brush, to start with.

Mr. GREGORY. I think there are also ways in which content developers creating content—excuse me—thank you—people creating content can also choose to add information to that data, so there are ways we can do this at the level of very minor modifications. There are also ways in which you could be tracking those desired usages, for example, using the C2PA standard.

So I think the more options we can give people that are standardized for understanding the choices they make about the information they consume, but also the information they place online, would be appropriate.

Senator HICKENLOOPER. Great. Thank you.

Ms. Espinel, what are some evidence-based steps that your member companies have been using to develop AI with the safety of consumers in mind, but in that notion of the AI Risk Management Framework?

Ms. ESPINEL. Thank you. So BSA members have—many of them have implemented very extensive risk management programs that include, as part of that, impact assessments. And I will talk a little bit about what they do, and how they can use evidence to make the kinds of determinations that you are talking about, to both increase transparency, but also reduce the risk of bias and discrimination.

So as an example, if a BSA member is acting as a developer of AI, they can assess the training data that they are using to ensure that it is representative of the community. And they can use the evidence that is gathered from that assessment to ensure that the risk of bias and discrimination is as low as possible.

That is certainly in line with the NIST AI—the risk management framework that was developed by NIST. But I would say, as important as the NIST Risk Management Framework is, and as much commendation I give to the Department of Commerce for coming up with it, we don't think it is sufficient to—we think it would be best if legislation required companies in high risk situations to be doing impact assessments and have internal risk management programs.

So yes, there is much that our companies have been doing. I think there is probably much that many companies have been doing, but we think in order to bring clarity and predictability to the system, and to ensure that use of artificial intelligence is as responsible as possible for Congress to require impact assessment, and require risk management programs, is essential.

Senator HICKENLOOPER. All right. Thank you very much. I am going to come back for more questions later.

Ranking Member Blackburn.

Senator BLACKBURN. Thank you so much. Ms. Espinel, I want to come to you first. Because I know that the EU is looking at implementing their AI Act later this year. I was recently over there, and working on the privacy issue, and holding some meetings, and I think they are a little bit frustrated that we haven't moved forward on some of these guidelines and governance that would help our innovators here in the country.

So talk for just a minute about how your members would navigate a patchwork system around the globe when it comes to AI and the guardrails that are going to be put around it?

Ms. ESPINEL. Thank you very much. So let me start off my thanking you. When you were in Brussels, you had an opportunity to visit us at BSA.

Senator BLACKBURN. Right.

Ms. ESPINEL. And visit many of our member companies. You led a fantastic roundtable there. So thank you very much for that. But also thank you for that question, because I think that is a very important issue.

So as you said the EU is moving forward with legislation. They are not the only country that is moving forward with legislation; governments around the world are moving forward with legislation as well. And I think one of the challenges, but an avoidable challenge, is if we end up in a situation where there is, as you said, an inconsistent patchwork of regulations around the world.

I think because there is so much policymaker focus around the world on artificial intelligence, as you said, in part, because of the launch of ChatGPT, there is a real opportunity right now to move towards a harmonized approach globally. And that may not include every country in the world, but I think it could include a lot.

And I think if the United States—I think the United States has a very important role to play there, in terms of moving a large number of countries to a harmonized approach.

Senator BLACKBURN. Should we be moving to set standards and lead this?

Ms. ESPINEL. In my opinion, yes.

Senator BLACKBURN. Okay.

Ms. ESPINEL. The United States is leader in innovation; we should be a leader here as well.

Senator BLACKBURN. Absolutely. Mr. Strayer, when you are working with your members, the National Data Privacy Law, how did they put importance on that before we move forward with some of these other components dealing with AI?

Mr. STRAYER. We believe a comprehensive national privacy law is absolutely critical, that ensures a lot of the issues that will come about, about data training sets, and other data that emerges from AI systems that are being used by businesses every day. So we pretty much support that—you know, acting quickly, we don't think those need to be done—but needs to be done first before moving on AI regulation, but we think both have to be done.

The thing I would say about standards is, U.S.-based companies and western companies, generally, are leading in developing stand-

ards through the International Standards Organization. They are working now on an AI management system standard. Those will be, hopefully, the bedrock for when EU finishes their legislation, the standards that should apply globally, but that has not been fully resolved by the European Union. But those standards should be to harmonize global standards for the future.

Senator BLACKBURN. Yeah. Mr. Gregory, do you have anything that you want to add on that?

Mr. GREGORY. I would note that one of the areas that the EU is focused on is labeling and disclosure of AI-based content. I think there is a real opportunity for the U.S. to lead on this, to come up with a standardized way that respect privacy, that presents information that is useful to consumers, and to set a standard there that is applicable as well.

Senator BLACKBURN. Yeah. Dr. Krishnan?

Dr. KRISHNAN. I think the missed AI RMF offers an opportunity here through what is called the NIST AI RMF profiles, and through the use of these profiles I think we could, with the appropriate standard-setting for high-risk applications, both on the data—input side, the data transparency side, as well as with model validation. We can actually come up with standards that then get adopted, because it is considerable support for AI RMF, both here at home and abroad.

Senator BLACKBURN. Thank you. Ms. Espinel, let me come back to you. I have got just a little bit of time left. I would like for you to just briefly talk about how your companies have worked to have policies that are transparent, that are interpretable, that are explainable, when it is—when you are dealing with AI?

And Mr. Strayer, I will come to you for the follow-on on that.

Ms. ESPINEL. So it is very important that we have a system that builds trust in AI, and transparency is clearly part of what is important in order to build trust. Let me give you just a few examples of ways that there could be transparency. One is to let consumers know if they are interacting with an AI service, such as the chatbot.

Another example would be, let consumers know if they—if the image that they are seeing has been generated by artificial intelligence. There is important work that is being done by other witnesses at this table, in terms of content authenticity, and letting consumers know if images have been altered or manipulated in some ways.

Those are all—those are just three examples. But I want to end by saying, in addition to transparency practices I do think it is very important that we have regulation, and that we have regulation that requires high-risk uses for companies that are developing or using AI to be doing impact assessments, to try to mitigate those risks.

Senator BLACKBURN. Mr. Strayer.

Mr. STRAYER. I would just add that companies are also setting up fact sheets for what they call model cards that explain the features and limitations, to talk about where the datasets came from, intended uses. So these are, you know, pretty fulsome explanations. It is important in the area of transparency to think about who the intended audience is, and for what purpose.

So is it a consumer, is it business as well on the chain? Is it for the deployer? So one should think about all those when they think about what requirements should be set in this area.

Senator BLACKBURN. Thank you.

Senator HICKENLOOPER. All right. Thank you.

Senator Moran.

STATEMENT OF HON. JERRY MORAN,
U.S. SENATOR FROM KANSAS

Senator MORAN. Chairman Hickenlooper, thank you. Thank you all for your presence and testimony today. It is annoying that we are here now on AI when we have been unsuccessful in reaching conclusions on data privacy legislation. It just seems like one issue piles up after another, both of huge significance.

Ms. Espinel, let me start with you. I want to talk about the NIST AI Risk Management Framework, launched after NDAA authorization in 2023. I am working on legislation, in fact, authored an amendment to NDAA this year, that would require federal agencies to apply the AI framework when using AI systems. An attempt to ensure that government acts responsibly in implementing AI systems, and in a manner that limits potential risk, not only to Americans and their data, but to governmental agencies and their missions.

Can you talk about, Ms. Espinel, about the implementation of policies based on NIST AI Risk Management Framework that can establish a baseline of good behavior when implementing artificial intelligence systems, which can actually unleash beneficial AI technologies instead of just hindering the development of AI?

Ms. ESPINEL. I would be delighted to. It is very important the NIST AI framework is flexible, it provides a roadmap for companies in terms of how they can put practices and policies in place, to responsibly develop and use AI. We support it being used by the U.S. Government, we support it being used in the context of procurement.

And I will close by saying, I think it is the place, that you kind of alluded to at the end, and as Ranking Member Blackburn raised, where the U.S. can show leadership. I think a model or an approach similar to the NIST AI Risk Framework, is one that could be usefully adopted by other countries as well; and so very supportive of that. Thank you.

Senator MORAN. Thank you. Mr. Strayer, I am going to go to you based upon your past history at—as Deputy Assistant Secretary for Cyber and International Communications and Information Policy at the State Department. That is a long title.

Mr. STRAYER. They give you long titles at the State Department.

Senator MORAN. Yes, sir. I hope the pay is commiserate.

[Laughter.]

Senator MORAN. Let me suggest the story to you of Huawei launching a phone last week, containing a suspected homegrown semiconductor that represents a leap forward in their ability to produce advanced microprocessors. Despite the efforts by the U.S. Department of Commerce to deprive that company of U.S. and partner technology to develop and manufacture advanced semiconductors, a lot of details yet to be known about that.

As part of that related effort to deny China the ability to develop these technologies, in August, President Biden issued an Executive Order limiting outbound investment of Chinese companies that develop advanced technologies, including AI. What are the national security implications for the U.S. if adversarial nations take a leading position in the advancement of artificial intelligence? Do you believe the U.S. can appropriately mitigate this risk through the current strategy of denying access to key technologies and investments? And what can we learn from our past and present efforts at restriction in the development of semiconductors?

Mr. STRAYER. Thanks Senator. That is quite a compound question I will try to do my best job to——

Senator MORAN. It goes with your title.

Mr. STRAYER. Touché. So first is, to maintain our leadership, we really also—we need to focus on running faster, that is how we innovate and make it to the next cycle of R&D in a position that puts us ahead of our adversaries, the United States, that is. So we need to continue to develop the best technology for all purposes, which will obviously benefit our military and national security uses for those.

On the defensive side, we have seen the October 7th, of last year, export control executive order regulation, and now this most recent outbound investment restriction.

We are still seeing those play out, there is open comment periods on these, they are trying to tighten them up. So we will see all those play out. There is a very important issue, though, with these. And that is, we really need to have a strong discussion with the private sector about how you can do the minimum amount of damage to future R&D and revenues for U.S.-based companies, while ensuring that none of these tools end up in the hands of adversaries, where we are going to use those for military purposes.

So really sharpen your focus on where it might be used for dual use or military purposes, while benefiting U.S. and Western companies to our asymmetric advantage, because they need to keep maintaining those revenues over the long term. So I think we need a stronger focus on working with the private sector to get that balance, right, that is stopping the military uses of other countries, and enhancing our own use and market competitive development of the technology.

Senator MORAN. Thank you. I have a long list of questions as well. But my time, at least at this moment, has expired.

Senator HICKENLOOPER. We will have a second round. Did you feel—did you get an answer on the restriction—is the restriction effective? Because I am not quite sure we all—you know, I would want to hear a little more about that?

Senator MORAN. As long as that is on your time, it is a great question, Chair.

[Laughter.]

Senator HICKENLOOPER. My time is your time.

Senator MORAN. Good.

Mr. STRAYER. So these restrictions are quite robust, in the case of export controls from the October 7th regulation that the Commerce Department issued on more advanced semiconductors. With regard to the outbound investment restrictions, they are starting

more narrow, and they are doing rulemaking through the Treasury Department on these. I think that is a smart way to start, and then start to expand if they need to beyond that.

The really key issue with these restrictions is that we don't want to force all technology—or key innovative technology to move outside the United States, so we need to bring allies and partners along with the United States, in the case of export—and the export controls for semiconductors, Japan, and others have come along, Taiwan, come along with the United States, the Netherlands, on the equipment as well.

That has not yet occurred on the outbound investment restriction, so I think one needs to think about how that is done in a multilateral way, so that we are not making United States a place where you don't do investments in these areas, and we are ceding that leadership to other—even Western countries.

Senator HICKENLOOPER. Isolating ourselves. Got it. Thank you.

Senator Klobuchar, we have remotely, for a few questions. I can't hear you.

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

Senator KLOBUCHAR. — Judiciary, so I truly appreciate you taking me remotely, and with many of the same issues. We are focused on many things with AI, and we should be from the security risk to the effects on innovation to, of course, the potential that we have to use the technology for good. And one of the things that we are really concerned about, which is off this topic a little, but it is just, to always keep in mind, as the experts that you are, is just the effect this is going to have on our democracy and our elections.

And so two things, one is that Senator Hawley, and Senator Collins, and Senator Coons, and I just introduced a bill in the last hour about deceptive AI-generated content. And that is the most extreme, right. That is the stuff where you have got people acting like they are the candidate when they are not, which is going to AI generated images acting like they are a candidate when they are not, which is going to create complete confusion, regardless of whether it is labeled or not.

And I thought it was really important to do this on a bipartisan basis, so I hope my colleagues will, and you can't get much more bipartisan than Hawley, and Collins, and Klobuchar, and Coons. And so I hope that my colleagues will look at this. It is about fraudulent AI-generated content in political ads.

The second thing that we are looking at, and for another class of AI-generated political ads, would be disclaimers, and watermarks, and the like, for the ones that don't meet the standard for the most extreme, deceptive, with of course, exception for satire, because we all like a lot of satire around here. Okay. So I am just going to just ask you about the watermark piece, because we just introduced the other bill, the disclosure piece.

Mr. Gregory, and Dr. Krishnan, do you agree that without giving people information to determine whether an image or a video is created via AI, the generative AI poses a risk to our ability to hold free and fair elections?

Mr. GREGORY. The evidence already suggests that this is a problem both in the U.S. and globally given the capacities of these tools, so yes. I also believe that election content is a first place where it is possible to start with both visible disclosure and particularly indirect disclosure by labeling a piece of content, and also providing metadata that could explain it. The key part would be to protect the capacity for satire, as you note, that is essential to protect, yes.

Senator KLOBUCHAR. Okay. Well, I really do appreciate that answer, and also the timeliness of it, given that we are in election season. We have already seen the use of it against some of my colleagues on both sides of the aisle. And so people are very aware. I think it also, by the way, extends, and I will get your answer, Dr. Krishnan, in writing, if that is okay, because I want to move on to another question.

I wanted to move into the AI risk assessment and mitigation issue. we know that these systems have the potential to impact individuals in many key areas, especially if it evaluates rental insurance applications. I am working with Senator Thune on a bill to require companies that develop and use AI to identify risk and implement procedures to mitigate risk. This involves Department of Commerce Oversight.

Ms. Espinel, do you agree that both developers and deployers of the AI systems bear responsibility to mitigate risk before they release the AI on the market?

Ms. ESPINEL. I believe that both developers and deployers, users of AI, should have obligations. I believe they both have responsibility. I would emphasize that I think the obligations that are put on them should differ, depending—so that it reflects what they do. So a developer of AI is going to have information about how the data—the data that was used and how the AI was trained.

To use an example of a deployer, a bank is going to have information about how loans were actually made, and whether or not loans were made in a way that was disproportionately negatively impacted a part of the community.

So 100 percent, I agree with you. And thank you for thinking about the distinction between developers and deployers, the fact that both should have obligations, and that those obligations and requirements should reflect what they do; the information that they are going to have about the AI system in question, and the different steps that each can and should take to identify and mitigate risk.

Senator KLOBUCHAR. Thank you. We have also seen intellectual property issues with AI, songs, and the like, copyright, only around half the states have laws, that give individuals control over these, their name, image, and voice.

Do you agree, Ms. Espinel, that we need stronger protections for the image, likeness, and voices of creators?

Ms. ESPINEL. I know there have been instances where AI—there has been AI-generated information or content that has pretended to be someone that it wasn't. That is clearly wrong, and should be stopped. I am thinking about writer publicity, as you point out, that is not something that exist consistently throughout the United States. And so I think thinking about solutions to that problem,

which is clearly wrong and needs to be addressed, is very important.

Senator KLOBUCHAR. And just one quick other question of you, Mr. Gregory. In addition to some of the copyright issues we are talking about, we also have journalism issues, and we have—Senator Kennedy and I have a bill that allows the companies to—the companies to have to negotiate with news organizations on the issue of their content. And in addition to newspapers, AI systems are trained on other content like lifestyle magazine, most of which were not compensated for that content.

Do you agree, Mr. Gregory, that there is more we need to do to ensure that content creators are fairly compensated for their contributions to AI models?

Mr. GREGORY. Yes. I think there needs to be stronger ways to understand which content is being ingested into AI models, and the decisions that are made around that. And I would particularly highlight that journalists already face significant pressures, including of course, that they are unable to detect AI generated medias; they face pressures both if their content is ingested, and also that they are on frontlines of defending the truth in the context we face now.

Senator KLOBUCHAR. Thank you very much. Appreciate it, all of you. And thank you to the Chairman and Ranking Member for having this hearing.

Senator HICKENLOOPER. Thank you, Senator. Appreciate that. We don't have the next video—senator to question, so I will weigh in—oops—there we have Senator Young, just in the nick of time.

STATEMENT OF HON. TODD YOUNG,
U.S. SENATOR FROM INDIANA

Senator YOUNG. Thank you, Chairman, for acknowledging my presence in this room, and for chairing this hearing. I thank our witnesses for being here. I will just dive in. I know you have been responding to a number of inquiries.

I will begin with the observation that artificial intelligence wasn't something that we invented yesterday; it has been around for decades now. In fact, for years we have seen AI technology in products and services across nearly every sector of our economy, and in a wide variety of use cases.

Analysis of each of these use cases, and concerns of an AI-enabled society should, in my view, start with the same litmus test. Does existing law address whatever potential vulnerability we are concerned about?

I have found through my interactions with a number of experts in this area, that existing law would address the vast majority of concerns that we have, not every one though.

We have to closely evaluate and carefully target areas where existing law doesn't address these vulnerabilities, that is why, of course, we are here today, to identify high risk use cases of AI, and discuss potential guardrails to minimize those risks. Recent advancements, in generative AI platforms, like ChatGPT, have raised concerns among my constituents, and many others, about a dystopian future. Straight out of science fiction, that could be right around the corner.

The truth is, nobody knows what future this incredible technology will usher in, and it is human nature to fear uncertainty. History shows innovations that democratize access to information, into media, the printing press, recorded sound, film, have been met with similar concerns, usually exaggerated concerns. But history also shows these innovations have brought great benefits to society, to national security, and the economy.

As we evaluate the need for any level of AI regulation it is important we don't lose sight of the many potential benefits that harnessing the power of AI presents. These benefits include: self-driving cars, medical advances, immersive technology, educational opportunities, and more. So I sort of want to get on record, a high-level perspective.

With that said, you are here today, and let us not focus on the unknowns but rather the knowns, the here and now risks. As we think about trust, transparency, and exploitability within AI, the goal is not to stifle growth, but rather to increase adoption, and innovation in the long term.

Ms. Espinel, can you briefly discuss two things: first, the important distinction between a developer and a deployer? And then second, how should Congress think about the role of transparency between businesses and consumers as opposed to transparency between businesses and government? And I ask that you answer these pretty tightly, if you could, so I can ask some follow-up questions. Thanks.

Ms. ESPINEL. Thank you. So developers and deployers, developers and users, developers of AI systems are the ones that are producing the AI system, they are creating the AI systems, and the deployers are the users. To give an example, a developer is a software company that is developing a speech recognition system, and a bank is using, a deployer, using an AI system to help make determinations about who should get loans.

Those are very distinct roles, and the developer and the deployer will know very different things about the AI system, both how it is being developed, and how it is being used. And because they know different things, they will be able to do very different things in terms of addressing and mitigating that risk. So as you were thinking about legislation, clearly distinguishing between developers and deployers is very—is critical, in order for the legislation to be effective and workable.

In terms of transparency, you also alluded to the fact that—or mentioned the fact that AI has been used for a long time, right. It has. It is also used in many different types of circumstances, and some of those are high risk, and some of them are lower risk. And it is our believe that in high risk situations, so for example, making a—a government making a determination about access to public benefits as an example; if there are consequential decisions that impact a person's rights, we believe there should be legislation requiring that an impact assessment be done, and that those risks be mitigated.

But there are also uses, as you have said, that have been around for a long time, that are relatively low risk. So reminding me when I send an email that I may have left off an attachment. Or one that has been quite popular lately, adjusting, you know, the back-

ground, if you were on a video conferencing call. Those are relatively low risks, and having impact assessment required in those cases, we believe would be overly burdensome, and not add a lot of value.

But where there are other consequential decisions, whether by—whether companies to consumers, or government to its citizens, we believe impact assessment should be required.

Senator YOUNG. Well, thank you. Does anyone else want to chime in on any of those points? Otherwise, I will turn to Dr. Krishnan. Okay.

Doctor, what are the clear high risk use cases, to your mind, for AI that Members of Congress should be thinking about right now?

Dr. KRISHNAN. The ones that come to mind immediately are autonomous vehicles, health care, hiring, recruiting, housing, you know, there are—important scarce resources are being allocated via AI, those would all represent where there is harm either to the individual, or to society if things didn't go well.

Senator YOUNG. Right, right. And so your bringing up the use case, I am not surprised by. And then you would probably acknowledge there some in the national security realm——

Dr. KRISHNAN. Without a doubt, yes.

Senator YOUNG. Okay. Okay. I guess the last thing I would like to ask is, stepping back, AI has of course garnered all sorts of attention over the last number of months. Is AI all that different, and I will ask Dr. Krishnan, from other major technological advances? Or is it just the latest shiny object that we are attentive to? Why is this so important, or perhaps you would just say, no, this is like every other technology?

Dr. KRISHNAN. At one level, it is like other technologies we have dealt with in terms of having this impact on the way work is done, tasks that are impacted. And others are special characteristics of the technology in terms of working with the kinds of modality, audio, video, text, things that we haven't typically seen has not been part of a technological capability.

Like you mentioned ChatGPT in your opening remarks, that kind of capability was not something that—at least the typical citizen was thinking that this was something that a computer could do. And so the difference I guess, is also in terms of how the technology is able to learn over time, with data. So there are some differences that are technical differences with regard to this technology, and then there are differences with regard to how to govern the use of this technology.

And that is why in my testimony I talk about data transparency, and model transparency, and having standards for high-risk applications, then also having this trust infrastructure, because you can't predict exactly how this technology is going to evolve, to ensure that we are able to capture vulnerabilities, deal with failures, and come up with solutions that can be disseminated.

Senator YOUNG. On the fly, right; the trust infrastructure?

Dr. KRISHNAN. Yes.

Senator YOUNG. I guess the——

Dr. KRISHNAN. It is like the search for AI is how I think about it. Like what we have done for cybersecurity, sir.

Senator YOUNG. Maybe I can pull on that thread, just ever briefly, and end—you can respond to what for me has been an observation. I will leave it to you to determine whether or not it has been an insight. But my conclusion is, we are going to need a lot more expertise on this technology, a lot more sophistication within government, in individual agencies, perhaps at the White House, so that on an ongoing basis, we can figure out how to apply existing statutes to emerging threats, or concerns, or challenges, or opportunities.

Dr. KRISHNAN. Mm-hmm.

Senator YOUNG. Or to flag when new legislative authorities may be needed. Is that your estimation? The human resources challenge within government?

Dr. KRISHNAN. Yes. And in industry as well, so I think a scholarship for a service-type program, for AI, would be very, very valuable.

Senator YOUNG. Thank you.

Senator HICKENLOOPER. That last point was worth the price of admission. Thank you, Senator.

And I couldn't agree more. I think that if you look at the—try to estimate the cost of government keeping up with the rate of change and the innovation that is going to be required, it is a staggering thought.

And I have a son in college, and all the kids that are in STEM are looking at high-paying jobs right out of school just to start without the experience to be able to help government keep pace, and that it is that the competition is going to just fuel greater intensity and greater inflation among the wages, which again, is a good thing for the kids, but hard for the success of government relations within the industry. Thank you.

I am going to go into the second round. I could do a third and a fourth round. I will probably try and let you out of here by 4:00, a little bit after.

Mr. Gregory, you recently co-chaired the Threats and Harms Task Force within the Coalition for Content Provenance and Authenticity, my staff tells me it is referred to as C2PA. C2PA refers to the provenance as the basic trustworthy facts about the origins of a piece of digital content. This could help users distinguish between human- and AI-generated content. It could reveal personal information about the creator of that content as well.

So the question I have is, how do you—how do we protect, how do you protect, how do we protect the privacy of content creators while being transparent with consumers about the content they see when they are online?

Mr. GREGORY. As the Senator notes, I co-chaired the Threats and Harms Task Force. I don't speak for the C2PA here, but I will make some observations about how we protect privacy in this context.

I think this is absolutely critical. A starting point is to recognize we are moving into a much more complex media ecosystem. So the idea of understanding how media is made, how it has evolved, right where the AI has come in, where the human element has come in, I think is going to become increasingly important.

When we think about that future, though, we need to make sure that these systems are not either accidentally or deliberately, perhaps by authoritarian governments who might adopt them, systems, which they can track people's activity. That is why when we start looking at these types of approaches, we have to start from the principle that they do not oblige personal information or identity to be part of it.

That is particularly easy with AI-generated content because really what matters with AI generation is "how", not "who", right, the AI was used. The "who" could be helpful, but it is not necessary, and it could be helpful to the wrong people. So when we start from that premise, I think that is very important as Congress looks at how to standardize this, and how to approach this, that they understand how we are going to reflect the evolving nature of media production in a way that protects privacy, doesn't oblige personally identifiable information, and will be usable worldwide.

Senator Blackburn's question earlier about how the U.S. can compete. We should be developing standards that can be applicable globally. That means they need to be accessible, privacy-protecting, and usable in a variety of contexts globally.

Senator HICKENLOOPER. Absolutely. Ms. Espinel, increasing the transparency of AI systems is one of the key vehicles by which we gain confidence and trust among the users. Without appropriate guardrails around the risks from AI, I think developers will struggle to compete in the U.S., and certainly internationally as well. So it is in our best interest to demonstrate leadership, and safe and responsible deployment. What transparency steps do you prioritize; do you think are most crucial in terms of gaining the trust of consumers?

Ms. ESPINEL. So what if I start off by saying that I think building trust, as you say, is important for our ability to compete. I think it is important for the U.S. economy, and it is obviously important in terms of protecting consumers. So I think that is an absolutely critical step.

Impact assessments are a tool that we think that organizations should be using whether—again, whether they are creating the AI or they are using the AI. If they are in a high-risk situation, if the AI is being used to make a consequential decision, then impact assessments are an accountability tool that should be required. And by requiring impact assessments, you will increase transparency.

Consumers need to have confidence that if AI is being used in a way that could have an impact on their rights, or have, you know, significant consequence for their life, that that AI is being vetted, and that it is being continuously monitored to be as safe, as secure, as nondiscriminatory as possible.

And so I would go back to saying having a requirement for impact assessments by developers or deployers in high-risk situations, having a strong national law from the United States, I think is very important in terms of protecting consumers and our economy.

And then going to your last point, I think it is also very important for the United States to have an affirmative model of effective legislation when we—when other countries are moving quickly to regulate. And I think having the United States be a leader in shap-

ing that conversation, and the global approach to responsible AI is critically important.

Senator HICKENLOOPER. And an affirmative model; what a concept.

Ms. ESPINEL. Mm-hmm.

Senator HICKENLOOPER. Dr. Krishnan, you have done a lot of research with consumers and social behavior within digital environments, so on that same subject, what information should be disclosed to consumers to establish trust in online services around AI?

Dr. KRISHNAN. Well, first and foremost, I think when you are interacting with an AI system I think you need to know that you are interacting with an AI system.

Senator HICKENLOOPER. Okay.

Dr. KRISHNAN. So disclosure; that is the first step. The second is if data is going to be collected by the — by, let us say, a ChatGPT or Bard during your interaction, you should be explicitly given the option of opting in for the purposes of saying, "Is my data then going to be used by the AI to further—for training purposes?" That, I think, we can learn from much of our prior work in privacy to apply it in this kind of context. So the opt-in.

And then the third, I think, is with regard to the trust interaction that individuals have. You know, to a large extent, individuals build trust based on their own experience, much as we talk about data transparency, model transparency. My interaction with this tool: Does it actually behave the way I expect it to? That actually builds trust over time. And I think it is a combination of these that will result in the intended—you know, what we would like to see as an outcome.

One quick additional point I want to make is, while we have been talking about NIST, RMF, and the like, I think it would be great to have demonstration projects for citizens to recognize the value that AI can actually bring to the table. ChatGPT was perhaps the first time that they got to interact with AI on the kind of scale that we thought they were. It would be great to see something like the Khan Academy's education products, things of that nature. It gives them a clear indication of the value that this brings. I think that would be very good too.

Senator HICKENLOOPER. Couldn't agree more. I am going to sneak in one last question before the Chair comes back. I don't know if you guys play bridge, but the Chair trumps every suit. To be clear.

Mr. Gregory, let me go to my—switch to my question for Mr. Strayer. The AI ecosystem can be generally viewed as those that develop AI and those that use AI. As we have heard data is nuanced—there is a lot of nuance around that; risk management principles should be tailored both to developers and users, employers, and certainly there is not going to be any one-size-fits-all; there is no silver bullet. How can we create an oversight and enforcement framework to make sure that we can hold bad actors accountable? You know, people that use the AI systems maliciously?

Mr. STRAYER. Well, on the true malicious actors out there, there is going to need to be, you know, law enforcement cooperation and also enforcement of some of our existing laws. When it comes to standard risk management, a number of the appropriate risk man-

agement tools are going to make the model more resilient, more robust, less susceptible to compromise and manipulation. So those are important steps.

The other thing just to keep in mind is, at these risk management steps, you know, there should be higher risk management for the highest risk use cases, and lesser requirements on something that is just doing predictive text in an email.

And then finally, also to think a little bit about how small businesses and those that might be just doing experimentation that aren't planning for commercial deployment might be required at a lower standard than those that are going to make massive commercial deployments of things.

Senator HICKENLOOPER. That is such a good point; the small business aspect gets overlooked so often.

I am going to have to go vote, but I am going to leave you in the very capable hands of Senator Cantwell, who, as I said earlier, really knows probably more about this issue than anybody else in the Senate. So I am not worried that you are going to get lost in the forest.

The CHAIR. Thank you, Chair Hickenlooper. And again, thank you to you and Senator Blackburn for holding this important hearing; and for all our witnesses participating in this; I am sure it has been a robust discussion on many fronts.

I wanted to go back to the particulars of what you all think we should do on the deepfake side, as we see technology being developed, and DARPA playing a pretty key role as it is today, in looking at deepfakes and deepfake information. What is it you think is the landscape of a Federal role in identifying? Some have described a system of a watermark, some have described a—you know, immediate information system similar to what the Amber Alerts are, or something of that nature.

What do you all see as the key tools for effectiveness in developing a system to respond to deepfakes? And we will just go right down the——

Ms. ESPINEL. So it is a very important issue. I think there is a lot of great work that is being done, some of it spearheaded by a BSA member company named Adobe that has been working on the Content Authenticity Initiative. And I think in terms of giving—I know a lot of that is focused on making sure that consumers have more accurate information that is truly easily accessible, that they could access, and use, and take into account about the generation of AI content, and about whether or not that content has been manipulated or altered in other ways.

But I also know that there are witnesses at this table that are devoting a great deal of their life and energy to that thought. So I am going to cede the mic to them.

Dr. KRISHNAN. Senator, first a broad comment about trust. I think trust is a system-level construct, so when you think about humans interacting with machines, machines interacting with machines, one needs to think about what are the ways in which we can enable trusted interactions, trusted transactions to engage—to happen between them.

Deepfakes, as an example, I think content labeling and detection tools to go along with content labeling is absolutely essential to

allow for individuals, so that when I am interacting with a piece of content, for me to know whether it was actually AI-produced, whether it is a deepfake, so to have that information.

Equally well, beyond the technology piece, you need education for individuals to know how to actually process this information so that they can arrive at the right outcome with regard to this interaction between human and machine. Similarly, you could also have machine-to-machine exchanges of data where you could have, you know—I produce a piece of video content, and I pass it on to another machine, that has to be a—this is where standards are important. This is where C2PA, the standard you heard about, combined with watermarking, could actually provide the trust infrastructure to address this deepfake problem.

The CHAIR. Okay. Mr. Gregory.

Mr. GREGORY. I believe there are a number of steps the Federal Government can take. The first is to have a strong understanding of the existing harms and impacts, and really be able to understand where to prioritize with groups who are impacted. That includes harm as we know already like nonconsensual sexual images, but also the growing number of scams.

The second area would be to focus on provenance and to come up with a standardized way for people to understand both AI provenance, and opt-in. Human-generated provenance.

The third would be to focus on detection. Detection is not a silver bullet; it is flawed, but its availability is still limited to the people who need it most on the frontlines of journalism, human rights, and democracy, so continued investment from DARPA, and others, to really resource and support in diverse circumstances.

I believe there is a space for legislation around some specific areas, such as non-consensual sexual images, AI-generated CSAM, and potentially political ads that could be taken. And I believe it is the role also to look ahead and understand that this continuing ease of generation of synthetic media means that we will get more and more personalized, and this will have an impact in spaces like social media and platforms. So we should look ahead to those dimensions and be ready to consider those.

The CHAIR. Okay. Mr. Strayer.

Mr. STRAYER. I won't repeat what has already been said, but two things on the technical side very much to emphasize the importance of having an open standard for provenance. And secondly on the social dimension, you know, digital literacy is going to be really important for these things to be implemented. So bringing it to other stakeholders that include the media platforms, consumers, on the digital literacy side, for how these tools will be implemented effectively.

The CHAIR. So who do you think should be in charge of this; anybody? Mr. Gregory, you look like you are going to volunteer here.

Mr. GREGORY. I am going to volunteer, but I am probably not the best placed. So I will note that I see good leadership from agencies like the FTC that have been doing good work to support consumers to date; so supporting existing agencies that are doing good work with the resourcing and the support. In terms of the legislative gaps, I am not well placed to observe where those should come from. In terms of the R&D, I think that is broad support that ideal-

ly also goes outside of DARPA to other research facilities; and facilities more broadly in the U.S.

Dr. KRISHNAN. In my testimony, I think with regard to the content being produced, I think Congress should require close-source and open-source models to actually create this watermarking label and a detection tool to go with this label. This is for images and video. Text is a huge issue as to what it is—because you could have deepfakes with regard to text as well. And I think research is needed there. So I think it is a combination of things, but I think Congress should take a leadership role.

Mr. STRAYER. I will just add. Congress obviously has a very important role to play. I also think that NIST is a place where, over time, we have seen them deal with some very difficult problems, come up with new profiles for addressing very specific challenges and developing standards that are universally accepted, through in this process. And so I think NIST has a key role to play here too.

The CHAIR. Well that is why in the original legislation that we did with the NAIAC was to establish, you know, getting everybody together and figure out what we think the U.S. Government's role and responsibility should be. And while they haven't finished, you know, all of their findings, they have certainly made a list of the directions and recommendations. And so I think they are a good place to look for on this issue as well, at least from a discussion perspective.

But today's hearing was about stimulating some input about the issues around that. And what you basically are saying is there is no fail-safe way to do this. It is going to need constant participation, both on the side of making sure there is not mistakes; it is one of the reasons why I support getting a big—privacy bill that establishes a hard line against discriminatory action because then you could always take that action, again when somebody has had substantial harm given by a direction.

I think the privacy framework we have already laid out to basically stop that kind of activity and protect people. We have heard a lot from the civil liberties community about this, about what you might see as online redlining, and you worry about something in the machine learning environment just putting that into a system, and then it being there for years and years without anybody even understanding that there was a discriminatory tactic against somebody.

And all of a sudden, you know, all of these people don't have the same kind of thing, alone, that they wanted. And so this is something we definitely want to have a forceful bright line, in my opinion, against, and say that if these kinds of activities do exist, that we will stop them, and that we have a strong law on the books to prevent them from happening.

What do you think on the collaboration level, from an international basis, as it relates to deepfakes and communication? Anybody given that thought about how that framework should operate?

Mr. STRAYER. Just to point out, one analogy of the past was there was a lot of challenge with extremist — violent extremist content online in the—roughly in the mid-2000s post-9/11, there was something formed called the Global Internet Forum to Counter Terrorism, and that was really the major platforms. But then many

other players came together to form practices and procedures for getting this extremist content off the internet.

And so some kind of multi-stakeholder group coming together to do this is probably one of the best ways we can see this addressed expeditiously, as the problem will grow very quickly as well.

The CHAIR. Didn't Interpol play a big role in the early days of the internet in trying to do a similar thing? Trying to police against pornography online, and catching, you know, bad actors who were perpetrating content?

Mr. STRAYER. Absolutely, yeah.

The CHAIR. And so that was a—where an international organization was working, and organizations working with them to try to police, I guess, or create standards or information for people to stop those activities?

Mr. STRAYER. Yeah. Sort of a Clearinghouse model, I think that is how they pursued.

The CHAIR. And do you think that was successful?

Mr. STRAYER. They were, I think, a big component of it. I think the United States shouldn't shy away from taking credit for a lot of work that it did, bilaterally, through the Department of Justice to educate foreign partners about the ways that they can address things like pornography that rise to that level that is criminal. So I think the United States has been a real leader in ensuring security and safety on the internet.

The CHAIR. Thank you. Mr. Gregory.

Mr. GREGORY. To add that one of the gaps that we see frequently, and we support local journalists who are trying to identify deepfakes as well as local civil society, as they don't have access to skills and resources. So looking at mechanisms to share skills, capacity, fellowship that would bring that expertise closer to the people who need it. The circumstance we see very frequently right now is people claiming that real content is AI-generated, and people being unable to prove it is real.

And that is corrosive in many contexts around the world, and a lot of that has to do with the lack of access to skills and resources; so thinking about opportunities for the U.S. Government to support that.

The Chair: So what would that be? Because now you are talking about a subject very near and dear to my heart; and that is the erosion of local journalism by the commoditization of advertising. And I would say the non-fair use of big companies not giving media their fair value for content. You are not really—you know, it is not your content to keep the advertising revenue when it is within your browser instead of going to *The Seattle Times* or some other website. So this is a problem. And we have to fix that as well.

But you are saying their job is, you know, truth, justice, and the American way. And how can they detect that if they can't do the kind of investigations? Is that your point?

Mr. GREGORY. Yes. That they don't have access to the tools that they need, and so as DARPA and others build tools, making sure they are accessible and relevant to journalism and others, it is skills so that those are available, and that could be facilitated through existing programs that provide skill sharing.

I agree with you. There is a larger context where this is but a small symptom of a broader challenge to journalism, where AI increases those challenges, as well as provides opportunities for journalists to use it.

The CHAIR. Well, we definitely heard that in Seattle at our Summit, that we already have a problem as it relates to keeping and saving local journalism, and I am very concerned about it, because we have existed as a country for hundreds of years with this kind of oversight to make sure that the process that we all participate in works and functions, and the issues are brought up.

And clearly, we are seeing places in the United States where the journalism has—you know, ceased to have a credible model, that is a financial model, and thus we have seen the rise of a lot of very unfortunate issues, including corruption, because there is no one there to cover and watch the day-to-day.

So it is a very interesting question you are posing beyond what we do as a government in detecting deepfakes. How do you bring the oversight to those whose job it is to do oversight?

Mr. GREGORY. And whose job will get even more complicated in the coming years with the growth of AI-generated content?

The CHAIR. Yeah. And so do you think that is about misinformation, or do you think it is bigger than just misinformation?

Mr. GREGORY. I believe it is a question of misinformation to some extent. It is a question of the easy capacity to create a volume of information that journalists have to triage and interpret. It is a question of that against a backdrop of lack of resources.

The CHAIR. Okay, and so what would you do about that?

Mr. GREGORY. In the U.S. context it is very hard to work out how to direct further resources towards local journalism. One option would be to consider, as we look at the way in which content is being ingested into AI models, is there any financial support to journalistic entities as they do that; this is obviously something that is being considered in the social media context in other countries. I don't know whether that would be a viable option to address local journalism's needs.

The CHAIR. So how exactly would it work?

Mr. GREGORY. I don't know the model that would work in our context. We have certainly seen other contexts, globally, where governments have looked for ways to finance journalism from social media.

The CHAIR. Yes.

Mr. GREGORY. But it is not a viable option here in the U.S.

The CHAIR. Well, okay, I like that—the phraseology should be, local journalism is financing their—these websites and their models. That is what is happening here.

Mr. GREGORY. Yes.

The CHAIR. And we just haven't been able to find the tools to claw that back. But if we have to go and look at this fair use issue, we will go back and look at it, because we are not going to keep going this direction. And AI is an accelerant. It is an accelerant on everything. The information age is putting challenges, and AI will accelerate that. But we have got to harness the things that we care about and make sure that we get them right, because we want the

innovation, but we also want these particular issues to be resolved. So we certainly, certainly in Seattle, have had that discussion.

Dr. KRISHNAN. Can I briefly comment on this?

The CHAIR. Yes. Yeah, go ahead.

Dr. KRISHNAN. So on the first part, with regard to the tools, I do think that the kind of infrastructure for trust that we have built up with information security with the CERT, with SISA for instance, that that kind of capability, if you built it for AI as well, which could be fairly quickly stood up with FFRDCs, that gives us the capacity even across countries to track deepfakes, even if they don't necessarily adhere to a content standard like C2PA, because I don't think any individual organization has that capacity.

But something like the CERT could have that capacity because it will span dot-mil, dot-com, dot-gov concerns, and this capability and expertise will reside in something like that. That is with regard to your first question, with regard to how do we manage and harmonize standards across countries.

With regard to the second point, I think it is spot on with regard to values on the one hand, the capacity to licenses copyrighted content, and then how do you actually assess—that is on the input side, so if you think of the AI models as taking input data from, say, *The Seattle Times,* or things of that nature, how they—how do they declare first that they are using this data, and then compensating *The Seattle Times* fairly for that use of that?

On the output side, the interesting question is, is it the case that *The Seattle Times* is getting more traffic from the ChatGPTs and the Googles of the world? Or is it the case that the revenue that should have come to *The Seattle Times* is really going to ChatGPT or Bard. I mean, the argument has been that because they provide that entry point into a content that they are actually directing traffic that otherwise would not have found you.

So I think that requires, I think, analysis and research of the traffic with regard to who is going where and who is directing what to these sites, because I think that gets at this revenue point that——

The CHAIR. Well, I am pretty sure about 25 percent of the traffic that is generated online, that big sites are getting from news organizations, are really revenue that belongs to news organizations.

Dr. KRISHNAN. Right.

The CHAIR. Regardless of the commoditization of advertising. It is still revenue that belongs to the newspapers. And so to my point about this, is that our report, that this committee, at least when we were the authors of a report, we found that local journalism was the trusted news source.

Dr. KRISHNAN. Mm-hmm.

The CHAIR. This is the point. And that you have many voices, that that is the ecosystem that keeps the trust. I mean somebody could go awry. But guess what? The rest of the ecosystem keeps that trust. So you know, I think *The Seattle Times* would say it is a very viable, identifiable source of trust. If you were creating information off of their historical database of all *Seattle Times* ever published stories, which is a very long time, that is probably some of the most trusted journalistic information you could ever get, because they had to be in that business, right? But anybody who

would then take that content and then who knows do what with it, you know, obviously is a very, very different equation.

So look, I think—I want to go back to the international point for a second because I do think you mentioned a lot of organizations. I am not sure everybody grasped, or maybe I didn't grasp everything you were saying about that. What do you think we should be—do you think the NAIAC should be working in coordination right now with international organizations to discuss what a framework looks like? Or are you thinking this is more siloed within organizations like national security issues versus consumer issues, versus other things?

Dr. KRISHNAN. So then NAIAC does have a group that Ms. Espinel leads, as a working group. The AI Futures Working Group that I lead with regard to this trust infrastructure point that I was making, we have been focused on that, but it does have international implications. But perhaps Ms. Espinel can speak more to it.

Ms. ESPINEL. So I do—I have the honor of chairing the International Working Group for the NAIAC Advisory Committee. I would be—there are conversations that we are having internally about ways that NAIAC, as a committee, could be helpful. And either in terms of making recommendations to the administration, which is our mandate, or perhaps NAIAC as a committee, and I would be—some of them I can't talk about publicly here, although I would be—I would be happy to have follow-up conversations.

I can tell you about one though that I think goes to what you are talking about, which is, I think we believe that it is very important as governments are thinking about what the right approach is to regulating AI, or trying to address some of the concerns that have been raised by artificial intelligence, to make sure that those conversations are happening not just with the United States, not just with the United States, and the EU, not just inside the G7, the OECD, but to try to have that be a broad-based conversation, including bringing in emerging economies that are—have not typically been as much a part of some of these discussions as I think should be the case.

And so I think if we are going to end up with solutions that are really effective, for example, in deepfakes, that is going to have to be a global initiative, and I think it will be stronger and more effective if it is—if those discussions are happening with a group of companies—with a group of countries that represent different perspectives.

So emerging economies are going to have slightly different benefits and challenges, they need to be part of that discussion and that is——

The CHAIR. Well, besides——

Ms. ESPINEL. Sorry. I am kind of probably overly passionate about it. So I feel like I have gone on a bit too long.

The CHAIR. No, no. The question I was trying to get at as people—listen, this committee passed this legislation, we created the NAIAC, we said, here is your responsibilities, we hope you have been thinking about this because we have given you a few years to do so. And so I was wondering if the current thinking was a divide over the complexity of dealing with national security kind of

deepfakes in, you know, commercial, and citizen issues on deepfake.

And whether they—you would reach some conclusion on the international side of, there is a lot to this, and a lot to communicate and coordinate, because obviously the World Wide Web is a big, you know, open system. So you could say the United States is doing this, but you need others to participate. But consumer issue is very different from how we deal with national security issues. And so did the—has the organization come to any conclusion on that?

Ms. ESPINEL. I think the short answer is no. Just not to be overly legalistic, but there are significant restrictions on what I am allowed to say in a public forum, and I don't want to—I want to be very careful not to cross any lines.

So I can tell you that I think there are conversations happening about national security and consumers, on the point—I feel like it is fine for me to say, on the point that you were talking about, I don't see there being a real challenge. I don't see there being a lack of consensus on national security versus consumer issues, and be able to engage internationally on that.

The CHAIR. Well, they are just different organizations within our government and I am pretty sure they are internationally. So it just makes it challenging.

Ms. ESPINEL. It makes it challenging. And so I think one of the—so this I could say, I will just say, in my capacity as BSA, one of the—you have, for example, the U.K. government is hosting a global summit in the beginning of November, and I think one of the challenges they face is, who—if you are going to have a global summit that is intended to address the safety of artificial intelligence, which is what the UK has announced, who are you going to have—who is going to be part of that summit, and how many issues can they address?

Because there are a myriad of challenges, and as you say, they are often addressed by different parts of government. Speaking just for the United—speaking just in the context of the United States, I think having effective coordination across the Federal Government, I think there is more that could be done there. And I think that would be very, very helpful because you don't want these issues to get siloed, you don't want agencies to be doing things that are duplicative or in conflict.

The CHAIR. Okay.

Dr. KRISHNAN. And I will reach out to your office, Senator, about the trust infrastructure point that I made. I am happy to provide additional information.

The CHAIR. Well, we all know that we have lots of international organizations that are working on coordination on lots of internet issues, as it is today. I think the question is, you know, what—has anybody with the NAIAC, come up with a framework; before we start having these kinds of big discussions? So anyway, we will get more information.

I want to turn it back over to Chair Hickenlooper. Thank you so much for, again, holding this very important hearing.

I see that our colleague from New Mexico is here. And so sorry to take up so much time; I thought I had a free opportunity while you were off voting. Thank you all to the panel too.

Ms. ESPINEL. Thank you. If I get to say briefly, I am excited to be here to testify in my capacity as CEO of BSA, but I would also be happy to follow up with you in my capacity as a member of the NAIAC Committee.

The CHAIR. Thank you.

Senator HICKENLOOPER. And didn't I tell you I was leaving you in good hands? I love coming in on the end of a discussion, saying, God, how did I miss that? And certainly the traditional news, the trust of traditional news, and how to make sure that they are fairly compensated for the costs, but I don't think any of us know any traditional news organization that is worth a fraction of what it was worth 15 years ago. Just the nature of what has happened.

I turn it over to the good Senator from New Mexico.

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

Senator LUJÁN. Thank you, Mr. Chairman. Thank you to you, and to Ms. Blackburn for this important hearing. And it is great to be able to listen to Chair Cantwell as well.

To all the members of the panel, thank you so much for being here today and offering your expertise.

One area I have spent a considerable amount of time during my time in the Senate on this issue surrounding broadband, I will say, as opposed to AI, is in the area of transparency and making things easier to understand.

And what I mean by that is, Senator Cassidy and I, we introduced something called the TL;DR, which was, Too Long; Didn't Read Act. And we all know what those agreements look like. Pages and pages of documents and, you know, before you can download or use something, you know, you go down a little box and it says, "Accept terms of agreement", and people click that and they move on and they don't realize what is there.

I was also proud to see and advocate during the Bipartisan Infrastructure Bill, something called Nutrition Labeling for Broadband, to make it easier for consumers to compare services across the way.

Now, the same type of transparency and easy-to-understand documentation is also needed so that consumers know where they are interacting with AI, know how their information is going to be used, and know when content is generated by AI. And I am working on legislation to require this type of transparency and disclosures and responsibilities.

Now, Mr. Gregory, you mentioned that threats from synthetic media and disinformation most impact those already at risk, like minority groups that face other forms of threats and discrimination. Mr. Gregory, yes or no, are deepfakes and AI-generated disinformation already being created in Spanish and other non-English languages?

Mr. GREGORY. Yes.

Senator LUJÁN. And, Mr. Gregory, yes or no, do you believe technology companies invest enough resources into making sure AI systems work equally well in non-English languages?

Mr. GREGORY. Systems are not as adequate in non-English languages, and there are less resources invested in making them applicable in non-English languages; and when it comes to deepfakes, many tools work less effectively for detection in less dominant languages outside of English.

Senator LUJÁN. And with that being said, Mr. Gregory, would the current system make more communities even more vulnerable in the U.S. and globally through the risk of synthetic media?

Mr. GREGORY. This is exactly the concern that my organization has spent the last five years working on, is the vulnerability of communities in the U.S. and globally to synthetic media, because of their exclusion from access to tools and support around it.

Senator LUJÁN. Well, thank you for your work in that space and bringing attention to it. I introduced a piece of legislation, called the LISTOS Act, this year to address the lack of non-English investment in multilingual, large language models. Any action Congress takes on AI transparency must protect our most marginalized communities, including non-English speakers. And for those that don't know what LISTOS means, it means "ready," but it is an acronym, so I credit the staff for coming up with an acronym that calls to action to ensure that we would all get ready.

Now, AI systems are useful in a huge variety of scenarios, from making financial decisions, to medical diagnosis, and content modernization, and I believe government should also utilize AI for improving constituents' access to government services. This just goes to show the broad application of what AI can do.

Now, Ms. Espinel, AI systems are used in different sectors and for different purposes, and these various systems can have different kinds of outputs. For example, we have predictive AI making best guesses, and generative AI creating content. I have a hard time keeping up with a lot of this stuff; but maybe the Insight Forum tomorrow will help a little for folks like myself, but understandably. Consumers and other users of these systems will have very different experiences interacting in these systems.

Now, my question to you stems from, if people even know they are interacting with an AI system, which is not a given, so under that premise, isn't it then necessary that any transparency and oversight requirements be specific to the sector or use case of the AI system?

Ms. ESPINEL. So you are absolutely correct that AI is used in many different ways, and has many different outputs, and has many different consequences. So I would say a couple of things. One is, I think, in terms of having transparency so that, for example, a consumer knows if they are interacting with an AI service like a chatbot, I think that is important.

But I would also say, to build on that, is that looking at the specific uses, perhaps, as opposed to looking at a sector, but looking to see whether or not that use is going to have a consequential decision on someone's life. Is it going to impact their ability to access government services, to access public benefits, to get a loan, to get education, to get a job?

And if it is, if it is going to have a consequential decision in someone's life, then I believe, we believe at BSA, that companies

should be required, by legislation, to do impact assessments to identify those risks, and then take steps to reduce those risks.

Senator LUJÁN. I appreciate that very much.

Now, Professor, under the thought of the explosion of generative AI, which we know has serious implications for the integrity of our information ecosystems; do you agree that Congress should require tech companies to label any AI-generated content with disclosures and metadata that includes information on how the content was created?

Dr. KRISHNAN. Thank you, Senator. Yes, they should. And I want to also thank you for your first remark about TL;DR and the Privacy Nutrition labels. This is something that is absolutely essential in this space. It creates something that is easy for people to understand. And as I understand it, Apple had introduced, two years ago, these labels that are privacy schematics that are easy to understand for people, and those need to be adapted for the purposes of what it is you were introducing in your opening remarks.

Senator LUJÁN. I appreciate that very much, Professor.

Now, Mr. Chairman, I have several other questions, but I see my time has expired. I have already exceeded it. I will submit them into the record. But I just want to thank you all for what you have been doing, what you are going to continue to do, those that are participating in one form or another with the forum tomorrow, and especially for the follow-up coming out of this as well.

And so, for those of you I did not speak to very much, or at all, I have some questions for you as well, and I will get them to you. So, I look forward to hearing back from you, and your expertise on the subject.

Thank you, Mr. Chairman.

Senator HICKENLOOPER. Thank you, Senator Luján.

And once again, I thank all of you. I am going to wrap it up, and unless we missed something urgent, speak now or forever hold your peace; although that is from a different ceremony. Today's discussion is important to help set the next steps for working with consumers to understand the benefits, and ultimately to trust AI.

Senators are going to be able to submit questions, not to Senator Luján, but all the senators, additional questions for the record until the hearing record closes on September 26.

We ask witnesses to provide responses to the Committee by October 10. And again, thank you very much for your time.

We are adjourned.

[Whereupon, at 4:30 p.m., the hearing was adjourned.]

A P P E N D I X

PREPARED STATEMENT OF HON. TED CRUZ, U.S. SENATOR FROM TEXAS

Thank you Chairman Hickenlooper and Ranking Member Blackburn. And thank you Chairwoman Cantwell for calling this hearing. This Committee should be at the forefront of discussions about artificial intelligence, or AI.

It is important to keep in mind that while AI is becoming more advanced, it is not new. It's been around for decades. You use AI every day without realizing it—for example, when you receive online shopping recommendations or have your text messages autocorrected. Beyond improving mundane tasks, AI is already transforming our world for the better. It's detecting cybersecurity threats to critical infrastructure; improving agricultural yield; and with advancements, potentially enhancing cancer detection and treatment. In these ways and more, AI has already vastly improved Americans' quality of life.

Congress would do well to learn from the past. This is far from the first time we've debated whether and how to regulate innovation. Take the Internet as an example. In the 1990s, the United States made a conscious, bipartisan decision to avoid heavy government intervention that might stunt the internet's growth, including bureaucratic organization under one agency head and influence over speech and content issues. The results speak for themselves: The U.S. now has the most successful Internet companies in the entire world—it isn't even a close contest.

With AI, I'm afraid that we are allowing history to repeat itself—only this time, we are following our European counterparts, who made a mistake with their early regulation of the internet. You can't read the news today without encountering Terminator-style fearmongering about AI building a weapon of mass destruction or developing sentience that will destroy humans.

Let's be clear: AI is computer code developed by humans. It is not a murderous robot. Humans are responsible for where and how AI is used.

Unfortunately, the Biden Administration and some of my colleagues in Congress have embraced doomsday AI scenarios as justification for expanded Federal regulation. Some of these proposals are extremely onerous: Licensing regimes, creating a new regulatory agency to police computer code, and mandatory, NEPA-style impact assessments before AI can be used. The fearmongering around AI has caused us to let our guard down to accept so-called "guardrails"—pseudo-regulations disguised as safety measures. These are often cumbersome and infeasible, especially for the startups so common in the tech sector.

I don't discount that there are some risks associated with the rapid development and deployment of AI. But we must be precise about what these risks are. I've heard concerns from my colleagues about misinformation, discrimination, and security. These certainly present challenges, but I have a hard time viewing them as existential risks, or even worthy of new regulation.

To me, the biggest existential risk we face is ourselves. At this point, Congress understands so little about AI that it will do more harm than good.

It is critical that the United States continue to lead in AI development—especially when allies such as the European Union are charging toward heavy-handed regulation.

Let me propose an alternative. Instead of riling fears and pausing AI development, let's pause before we regulate.

We can start by fully assessing the existing regulatory landscape before burdening job-creating businesses—especially startups—with new legal obligations. Many of our existing laws already apply to how AI systems are used. For example, the Fair Credit Reporting Act protects consumer information with reporting agencies and the Civil Rights Act of 1964 prohibits discrimination. We should faithfully enforce these laws, as written, without overstepping.

The FTC's investigation of OpenAI is a clear abuse of authority. As I wrote to Chairwoman Khan this week, fearmongering and fanciful speculation do not justify enforcement action against Americans creating new AI tools. The FTC's unprece-

dented and aggressive policing of AI would undoubtedly require statutory authority from Congress.

Leading the AI race is also important for national security. If we stifle innovation, we may enable adversaries like China to out-innovate us. I've been cautioning against ceding leadership on AI development to China since 2016, when I held Congress's first-ever hearing on AI. We cannot enact a regulatory regime that slows down innovation and lets China get ahead of us.

Think about if we had let fear get the best of us with other technological developments. Panics about new technology have occurred throughout history—and the panics have generally not borne out. There was widespread fear and calls to regulate around the advent of innovations such as automobiles, recorded sound, typewriters, and weaving machines. Every time, after the hysteria died down, we adapted and allowed technology to improve our lives, spur economic growth, and create new jobs.

The same opportunity exists today with AI. Let's not squander it.

R STREET INSTITUTE
Washington, DC, September 12, 2023

Hon. JOHN HICKENLOOPER,
Chair,
Subcommittee on Consumer Protection,
Product Safety, and Data Security,
Commerce, Science, & Transportation
Committee,
United States Senate,
Washington, DC.

Hon. MARSHA BLACKBURN,
Ranking Member,
Subcommittee on Consumer Protection,
Product Safety, and Data Security,
Commerce, Science, & Transportation
Committee,
United States Senate,
Washington, DC.

Dear Chairman Hickenlooper, Ranking Member Blackburn and members of the Subcommittee:

Thank you for your decision to hold a hearing on September 12, 2023 titled, "The Need for Transparency in Artificial Intelligence." My name is Adam Thierer and I am a senior fellow at the R Street Institute. I also recently served as a commissioner on the U.S. Chamber of Commerce's Commission on Artificial Intelligence Competitiveness, Inclusion, and Innovation.[1]

Artificial Intelligence (AI) technologies are already all around us and they are helping make our lives better in many ways. But the potential for algorithmic systems is even greater and these technologies also have important ramifications for our country's global competitive standing and geopolitical security.

The United States must reject the regulatory approaches being advanced by China, Europe and other nations, which are mostly rooted in a top-down, command-and-control approach to AI systems.

Instead, America's approach to technological governance must continue to be agile and adaptive because there is no one-size-fits-all approach to AI that can preemptively plan for the challenges that we will face even a short time from now.[2]

At this early stage of AI's development, government's role should be focused on helping developers work toward consensus best practices on an ongoing basis.[3] In this regard, the National Institute of Standards and Technology (NIST) has taken crucial steps with its *AI Risk Management Framework,* which is meant to help AI developers better understand how to identify and address various types of

potential algorithmic risk.[4] NIST notes it "is designed to address new risks as they emerge" instead of attempting to itemize them all in advance.[5] "This flexibility is particularly important where impacts are not easily foreseeable and applications are evolving," the agency explains.[6]

---

[1] U.S. Chamber of Commerce, *Commission on Artificial Intelligence Competitiveness, Inclusion, and Innovation: Report and Recommendations* (March 2023). *https://www.uschamber.com/technology/artificial-intelligence-commission-report.*

[2] Adam Thierer, "Getting AI Innovation Culture Right," R Street Institute Policy Study No. 281 (March 2023). *https://www.rstreet.org/research/getting-ai-innovation-culture-right.*

[3] Lawrence E. Strickling and Jonah Force Hill, "Multi-stakeholder Internet governance: successes and opportunities," *Journal of Cyber Policy* 2:3 (2017), pp. 298–99. *https://www.tandfonline.com/doi/abs/10.1080/23738871.2017.1404619.*

[4] National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework (AI RMF 1.0),* U.S. Department of Commerce, January 2023. *https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.*

[5] Ibid., p.4

[6] Ibid.

While it is always important to consider the dangers that new technologies could pose, extreme regulatory solutions are not warranted. Safety considerations are vital, but there is an equally compelling public interest in ensuring that algorithmic innovations are developed and made widely available to society.

Toward that end, AI governance should be risk-based and focus on system outcomes, instead of being preoccupied with system inputs or design.[7] In other words, policy should concern itself more with actual algorithmic performance, not the underlying processes. Transparency and explainability are important values that government can encourage, but these concepts must not be mandated in a rigid, overly prescriptive fashion.[8]

Algorithmic systems evolve at a very rapid pace and undergo constant iteration, with some systems being updated on a weekly or even daily basis. If policy is based on making AI perfectly transparent or explainable before anything launches, then innovation will suffer because of endless bureaucratic delays and paperwork compliance burdens. Society cannot wait years or even months for regulators to eventually get around to formally signing off on mandated algorithmic audits or impact assessments, many of which would be obsolete before they were completed.

Converting audits into a formal regulatory process would also create several veto points that opponents of AI advancement could use to slow progress in the field. AI innovation would likely grind to a halt in the face of lengthy delays, paperwork burdens and significant compliance costs. Algorithmic auditing will always be an inexact science because of the inherent subjectivity of the values being considered.

Auditing algorithms is not like auditing an accounting ledger, where the numbers either add up or don't. When evaluating algorithms, there are no clear metrics that can quantify the scientifically correct amount of privacy, safety or security in a given system.

This means that legislatively mandated algorithmic auditing or explainability requirements could also give rise to the problem of significant political meddling in speech platforms powered by algorithms, which would raise free speech concerns. Mandated AI transparency or explainability could also create some intellectual property problems if trade secrets were revealed in the process.

This is why it is essential that America's AI governance regime be more flexible, bottom-up, and driven by best practices and standards that evolve over time.[9] Beyond encouraging the private sector to continuously refine best practices and ethical guidelines for algorithmic technologies, government can utilize the vast array of laws and regulations that already exist before adding new regulatory mandates. The courts and our common law system stand ready to address novel risks that are unforeseeable in advance. Many agencies are also moving aggressively to consider how they might regulate AI systems that touch their fields. Using various existing regulatory tools and powers like product recall authority and unfair and deceptive practices law, agencies can already address algorithmic harms that are proven. We should not be adding another huge Federal bureaucracy or burdensome licensing mandates to the mix until we have exhausted these other existing solutions.[10]

The United States must create a positive innovation culture if it hopes to prosper economically and ensure a safer, more secure technological base. Policymakers must not try to micro-manage the future or pre-determine market outcomes. It is essential that we strike the right policy balance as our Nation faces serious competition from China and other nations who are looking to counter America's early lead in computational systems and data-driven digital technologies.

Sincerely,

/S/ ADAM THIERER,
Senior Fellow,
R Street Institute.

---

[7] Adam Thierer, "The Most Important Principle for AI Regulation," R Street Institute *Real Solutions,* June 21, 2023. *https://www.rstreet.org/commentary/the-most-important-principle-for-ai-regulation.*

[8] Comments of Adam Thierer, R Street Institute to the National Telecommunications and Information Administration (NTIA) on "AI Accountability Policy," June 12, 2023. *https://www.rstreet.org/outreach/comments-of-the-r-street-institute-to-the-national-telecommunications-and-information-administration-ntia-on-ai-accountability-policy.*

[9] Adam Thierer, "Flexible, Pro-Innovation Governance Strategies for Artificial Intelligence," R Street Institute Policy Study No. 283 (April 2023). *https://www.rstreet.org/research/flexible-pro-innovation-governance-strategies-for-artificial-intelligence.*

[10] Neil Chilson and Adam Thierer, "The Problem with AI Licensing & an 'FDA for Algorithms,'" Federalist Society Blog, June 5, 2023. *https://fedsoc.org/commentary/fedsoc-blog/the-problem-with-ai-licensing-an-fda-for-algorithms.*

PREPARED STATEMENT OF HODAN OMAAR, SENIOR POLICY ANALYST, INFORMATION TECHNOLOGY AND INNOVATION FOUNDATION (ITIF)

Chairman Hickenlooper, ranking Member Blackburn, and members of the subcommittee, we appreciate the opportunity to share with you our thoughts on crafting policies to increase transparency in artificial intelligence (AI) technologies for consumers. ITIF is a nonpartisan think tank whose mission is to formulate and promote public policies to advance technological innovation and productivity. In this statement, we offer three considerations policymakers should keep in mind to ensure consumers are protected from harm:

1. While policymakers should encourage companies to adopt the NIST risk management framework, they should recognize that it is not a silver bullet for trustworthy AI. There are a variety of technical and procedural controls companies can employ to mitigate harm and policymakers should encourage companies to explore the full gamut of mechanisms to find those most contextually relevant.
2. Because increasing AI transparency can make some systems less accurate and effective, policymakers should fund research to better understand this tradeoff and evaluate policies for transparency against the impact on system accuracy.
3. Policymakers should hold AI systems to the same standard as human decisions, which are not always transparent.
4. Policymakers should direct NIST to support work on content provenance mechanisms, which are techniques that help users establish the origin and source of content (both AI-generated and human-generated), rather than create policies that simply require systems to disclose when output is AI-generated.

AI offers significant societal and economic benefits in a wide variety of sectors. The biggest risk to consumers is that the myriad opportunities AI offers will not be translated into all the areas where they can make a positive difference in people's lives.

However, there are several other areas of risk to consumers from businesses using AI. One is the creation of unsafe AI products and services, such as a company putting an AI chatbot that advises users to do dangerous things on the market. Another is the use of AI to deceive or manipulate unsuspecting consumers, such as a company using AI to create and spread fake reviews about their goods or services, which ITIF's Center for Data Innovation explores in its 2022 report "How Policymakers Can Thwart the Rise of Fake Reviews."[1] A third is the use of AI to commit crimes that harm consumers, such as using AI to support cyberattacks that steal their sensitive information. While there are other applications of AI that interact with consumers, such as the use of AI to make lending or credit decisions or AI used in employment decisions, we note that these are not in the scope of the subcommittee and therefore keep our comments focused on those that are.

1. While policymakers should encourage companies to adopt the NIST risk management framework, they should recognize that it is not a silver bullet for trustworthy AI. There are a variety of technical and procedural controls companies can employ to mitigate harm and policymakers should encourage companies to explore the full gamut of mechanisms to find those most contextually relevant.

Chairman Hickenlooper and Ranking Member Blackburn are right to state in their recent letter to technology companies that the National Institute of Standards and Technology AI Risk Management Framework (NIST AI RMF)—a framework that helps companies identify and mitigate potential risks from AI—can help protect consumers from harm and encourage companies to responsibly develop and use AI.[2] However, it is important to note that many facets of trustworthy AI cannot easily be translated into objective, concrete metrics and technical standards alone are not a silver bullet for trustworthy AI.

For instance, ensuring AI systems are robust and secure is one important element of creating trustworthy AI that protects consumers, and yes, one can employ audits to check how prevalent algorithmic errors are. There are various types of error-analysis techniques to check for algorithmic error, including manual review, variance

[1] Morgan Stevens and Daniel Castro, "How Policymakers Can Thwart the Rise of Fake Reviews," (Center for Data Innovation, September 2022), *https://datainnovation.org/2022/09/how-policymakers-can-thwart-the-rise-of-fake-reviews/*.

[2] "Hickenlooper, Blackburn Call on Tech Companies to Lead Responsible AI Use," press release, Apr 19, 2023, *https://www.hickenlooper.senate.gov/press_releases/hickenlooper-blackburn-call-on-tech-companies-to-lead-responsible-ai-use/*.

analysis (which involves analyzing discrepancies between actual and planned behavior), and bias analysis (which provides quantitative estimates of when, where, and why systematic errors occur, as well as the scope of these errors).

However, other facets of trustworthy AI, such as ensuring these systems are fair or unbiased, are subjective and cannot be reduced to fixed functions.[3] To see why, consider two e-commerce platforms that use AI algorithms to recommend products to their users. One platform employs an AI system with an objective function to recommend products solely based on customer preferences and purchase history, aiming to provide personalized recommendations without taking into account the price of the products. The other platform uses an AI system with an objective function that considers both customer preferences and product prices, trying to recommend products that not only match user preferences but also fall within the user's budget. Assume both AI systems are designed to be error-free. Even if both AI systems are functioning perfectly, they may have different suggestions for consumers from different socioeconomic backgrounds. Defining which system is more "fair" in this context can be complex, as fairness might involve considerations of affordability, accessibility, and equal opportunity to access desirable products.

This example demonstrates that achieving fairness in consumer product recommendations can be multifaceted and context-specific. Fairness may not have a one-size-fits-all definition. Rather than pursuing technical standards alone, policymakers should be pursuing the principle of algorithmic accountability. As the Center for Data Innovation explains in its 2018 report "How Policymakers Can Foster Algorithmic Accountability," this principle states that an algorithmic system should employ a variety of controls to ensure the operator can verify algorithms work in accordance with its intentions and identify and rectify harmful outcomes.[4] When an algorithm causes harm, regulators should use the principle of algorithmic accountability to evaluate whether the operator can demonstrate that, in deploying the algorithm, the operator was not acting with intent to harm or with negligence, and to determine if an operator acted responsibly in its efforts to minimize harms from the use of its algorithm. This assessment should guide their determination of whether, and to what degree, the algorithm's operator should be sanctioned.

Regulators should use a sliding scale of enforcement actions against companies that cause harm through their use of algorithms, with unintentional and harmless actions eliciting little or no penalty while intentional and harmful actions are punished more severely.

Defining algorithmic accountability in this way also gives operators an incentive to protect consumers from harm and the flexibility to manage their regulatory risk exposure without hampering their ability to innovate. This approach would effectively guard against algorithms producing harmful outcomes, without subjecting the public-and private-sector organizations that use the algorithms to overly burdensome regulations that limit the benefits algorithms can offer.

2. Because increasing AI transparency can make some systems less accurate, policymakers should fund research to better understand this tradeoff and evaluate policies for transparency against the impact on system accuracy.

One of the core tenets of transparent AI people cite is explainability. Explainable AI systems are those that can articulate the rationale for a given result to a query. Explanations can help users make sense of the output of algorithms. Explanations may be useful in certain contexts, such as to discover how an algorithm works. Explanations can reveal whether an algorithmic model correctly makes decisions based on reasonable criteria rather than random artifacts from the training data or small perturbations in the input data.[5]

However, it is well-documented that there is often a trade-off between explainability and accuracy. As a 2020 paper led by NIST researcher explains "typically, there is a tradeoff between AI/ML accuracy and explainability: the most accurate meth-

---

[3] Rediet Abebe, Jon Kleinberg, & S. Matthew Weinberg, "Subsidy Allocations in the Presence of Income Shocks," Proceedings of the AAAI Conference on Artificial Intelligence (2020): 34(05), 7032–7039, *https://doi.org/10.1609/aaai.v34i05.6188.*

[4] Joshua New and Daniel Castro, "How Policymakers Can Foster Algorithmic Accountability" (Center for Data Innovation, May 2018), *http://www2.datainnovation.org/2018-algorithmic-accountability.pdf.*

[5] "AI Foundational Research—Explainability", NIST, *https://www.nist.gov/topics/artificial intelligence/ai-foundational-research-explainability.*

ods, such as convolutional neural nets (CNNs), provide no explanations, while more understandable methods, such as rule-based systems, tend to be less accurate."[6]

Policymakers should seek to understand the extent to which this is true for applications that impact consumers and how they can implement policies for increased transparency in a way that does not harm system accuracy. A 2022 paper called "The Non-linear Nature of the Cost of Comprehensibility," published in the *Journal of Big Data* notes that "while there has been a lot of talk about this trade-off, there is no systematic study that assesses to what extent it exists, how often it occurs, and for what types of datasets."[7] It could be the case that high-risk consumer-facing AI applications are more likely to become less accurate if they were made to be more explainable, or it might not. More research would help answer this question. Furthermore, if policymakers want to increase transparency for certain high-risk scenarios, they should also fund research into methods that might limit the impact on system accuracy.

3. Policymakers should hold AI systems to the same standard as human decisions, which are not always transparent.

Policymakers should be careful of holding AI systems to a higher standard than they do for humans or other technologies and products on the market. This is a mistake the European Commission is making with its AI Act. The EU's original proposal contains impractical requirements such as "error-free" data sets and impossible interpretability requirements that human minds are not held to when making analogous decisions.[8] Policymakers should recognize that no technology is risk-free; the risk for AI systems should be comparable to what the government allows for other products on the market.

More broadly, targeting only high-risk decision-making with AI, rather than all high-risk decision-making, is counterproductive. If a certain decision carries a high risk of harming consumers it should make no difference whether an algorithm or a person makes that decision. For example, if it is harmful to deceive consumers by creating fake reviews, enforcement action should be proportional, regardless of whether a human or an AI system was used to create them. To hold algorithmic decisions to a higher standard than human decisions implies that automated decisions are inherently less trustworthy or more dangerous than human ones, which is not the case. This would only serve to stigmatize and discourage AI use, which would reduce its beneficial social and economic impact.

4. Policymakers should direct NIST to support work on content provenance mechanisms, which are techniques that help users establish the origin and source of content (both AI-generated and human-generated), rather than create policies that simply require systems to disclose when output is AI-generated.

Some policymakers advocate for policies mandating that generative AI systems, such as those used in customer service, social media, or educational tools, must include notices in their output, informing users that they are interacting with an AI system rather than a human.

However, mandatory disclosure requirements may not always be practical or desirable. Many AI applications aim to replicate human capabilities, whether by crafting human-like e-mails or simulating lifelike customer service interactions. In such cases, labeling content as AI-generated could undermine the very purpose for which consumers use these systems.

Instead, policymakers should support content provenance mechanisms. Content provenance mechanisms are techniques used to trace and establish the origin or source of digital content, whether it's text, images, videos, or any other form of data. For example, one technique is to embed secure metadata within digital files to provide information about the author, creation date, location, and other relevant details. Metadata helps users trace the origin of the content they interact with, whether it's AI-generated, human-made, or a hybrid of both. This approach provides transparency without mandating a disclosure that might compromise the utility of AI

---

[6] D. Richard Kuhn et al, "Combinatorial Methods for Explainable AI," October 2020, *Preprint: 9th International Workshop on Combinatorial Testing (IWCT 20), https://csrc.nist.gov/CSRC/media/Projects/automated-combinatorial-testing-for-software/documents/xai-iwct-short-preprint.pdf*

[7] Sofie Goethals et al, "The Non-linear Nature of the Cost of Comprehensibility," March 7, 2022, *Journal of Big Data, https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00579-2.*

[8] Patrick Grady and Kir Nuthi, "The EU Should Learn From How the UK Regulates AI to Stay Competitive," (Center for Data Innovation, April 2023), *https://datainnovation.org/2023/04/the-eu-should-learn-from-how-the-uk-regulates-ai-to-stay-competitive/*.

systems. It also addresses concerns about the proliferation of misinformation on social networks by allowing users to verify the source of the content they encounter.

The private sector is already conducting work in developing tools and research for content provenance, such as the industry-led Coalition for Content Provenance and Authenticity (C2PA), an initiative that is developing standards and technologies for verifying the authenticity and provenance of digital media content to combat misinformation, establish trust, and increase transparency. The subcommittee should encourage and fund NIST to support and bolster such work.

In conclusion, we appreciate the opportunity to provide our insights on enhancing AI transparency for consumers. Transparency can play a valuable role in achieving algorithmic accountability for some applications and we encourage the subcommittee to support research into when and how this mechanism can be used to support greater use of AI for consumers.

PREPARED STATEMENT OF JENNIFER HUDDLESTON, RESEARCH FELLOW, CATO INSTITUTE

Chair Hickenlooper, Ranking Member Blackburn, and distinguished members of the Consumer Protection, Product Safety, and Data Security Subcommittee.

My name is Jennifer Huddleston, and I am a technology policy research fellow at the Cato Institute. My research focuses primarily on the intersection of law and technology, including issues related to the governance of emerging technologies such as artificial intelligence (AI). I welcome the opportunity to submit a statement regarding the potential effects of government requirements around transparency in artificial technology.

In this statement for the record, I will focus on two key points:

- AI is a general-purpose technology that is already frequently in use by the average consumer and a wide array of industries. Overly broad regulations are likely to have unintended consequences and impact numerous existing products.
- The U.S. policy towards AI should remain flexible and responsive to specific harms or risks and seek to continue the light touch approach that has encouraged innovative activity and its accompanying benefits. Consumers and innovators, not government, are ultimately the best at deciding what applications of a new technology are the most beneficial to consumers.

The most basic definition of AI is a computer or robot that can perform tasks associated with human intelligence and discernment. Most of us have been encountering AI much longer than we realize in tools like autocorrect, autocomplete, chatbots, and translation software. While generative AI has garnered recent attention, the definitions of AI typically found in most policy proposals would impact far more than the use of ChatGPT.

The result is a proposal to require transparency that a product uses AI, which would mandate disclosures on many already common products like search engines, spellchecks, and voice assistants on smart phones without any underlying change to the product. Not only would this require compliance costs for these existing products, but it also impacts the value of the underlying transparency. If nearly every product requires a warning that it uses AI at some level in its processes, then such a transparency requirement becomes nearly meaningless to the consumer who is fatigued from seeing the same constant disclosure. Furthermore, such general labels fail to provide meaningful information that consumers can understand and act upon if desired. The government should not be in the business of being a user interface designer as the best ways to communicate such information will vary from product to product and use case to use case.

Because AI is a broad general-purpose technology, one-size-fits-all regulations are likely a poor fit. Building on the success of past light touch approaches that refrained from a precautionary approach to technologies including the internet, policymakers should resist the urge to engage in top-down rulemaking that attempts to predict every best-and worst-case scenario and accidentally limits the use of technology. Industrial policy that seeks to direct technological development in only one way may miss the creative uses that entrepreneurs seeking to respond to consumer demands would naturally find. For this reason, policymakers should ensure regulations are carefully targeted at specific harms or applications that would be certain or highly likely to be catastrophic or irreversible instead of broad general-purpose regulations.

Instead of a top-down approach, government should also consider the ways that it can work with innovators and consumers to resolve concerns that may require a degree of certainty, but for which static regulation is likely to be problematic. This

should include both looking at the possibility of removing potential barriers as well as supporting the development of industry best practices as appropriate. If necessary, these best practices could be more formalized to address concerns around specific harms or create legal certainty. As I discussed in comments to the NTIA in June 2023:

> "Soft law tools—such as multi-stakeholder working groups to develop best practices—may help identify appropriate limits on certain applications while also providing continued flexibility during periods of rapid development. These soft law tools also support the interactions of various interest groups and do not presume that a regulatory outcome is needed. In other cases, they may identify areas where deregulation is needed to remove outdated law or where hard law is needed to respond to harms or create legal certainty around practices. This also provides opportunities for innovators to learn of regulators and society's concerns and provide solutions that may alleviate the sense of unease while still encouraging beneficial and flexible uses of a technology.
>
> Soft law tools and best practices can be formalized in a way that provides opportunities for transparency and information sharing. In creating voluntary standards, best practices and other soft law tools can also address areas where direct harm is less clear, but a variety of concerns exist." [1]

Innovation is often disruptive and can bring with it a sense of unease and even fear. While such uncertainty and fear is currently seen with regards to AI, previous advances in information technologies and media have had similar concerns around many issues, including trust of information.[2] Societal norms often develop around the appropriate uses of technology and allow for a more flexible and responsive approach to concerns than a static law or regulation. Policymakers should build on the success of past light touch regulatory approaches that have made the U.S. a leader in technological innovation and narrowly tailor necessary interventions to respond to otherwise unaddressed harms and specific uses or applications of the technology.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. BRIAN SCHATZ TO VICTORIA ESPINEL

Question 1. Do people have the right to know when they are viewing AI-generated content or when they are interacting with an AI-chatbot?

Answer. Transparency is an important part of building trust in AI systems, and it can be realized in a range of different ways. BSA believes that companies that use AI systems to interact with individuals should inform those individuals that they are interacting with an AI system (such as a chatbot). BSA also supports the development of technical measures that can help users identify when an image is authentic.

Question 2. What are the dangers if individuals do not have a clear understanding that they are viewing or interacting with AI-generated content?

Answer. This is an important issue because deceptive content, such as deepfakes, can be indistinguishable from real content—with significant consequences. For example, deepfakes can lead people to believe that individuals have said or done things that they did not say or do. This can be especially harmful in scenarios, such as political campaigns, where it can spur confusion and misinformation. One concrete step we can take is to help consumers understand when the content they are seeing is real, which is critical to confronting disinformation. For example, the Content Authentication Initiative (CAI) and the Coalition for Content Provenance and Authenticity (C2PA) have been doing important work in this area to develop open standards that can create a stamp of authenticity.

Question 3. Do people have the right to know what model created the AI-generated content they are viewing or interacting with? Should generative AI systems affix information to AI-generated content reflecting that the content is generated by AI? Should that information be permanently affixed to the extent doing so is technically feasible?

Answer. Promoting transparency in AI systems is an important goal, but the type of information that is relevant to consumers will vary greatly depending on the type of AI system the consumer is interacting with and the purpose of that AI system. In many circumstances, identifying the model creator may not meaningfully further consumer transparency, because consumers may be more focused on other informa-

---

[1] *https://www.cato.org/public-comments/public-comment-regarding-ai-accountability-policy*
[2] *https://www.techdirt.com/2019/02/28/deception-trust-deep-look-deep-fakes/*

tion, such as a disclosure that the consumer is interacting with an AI chatbot. Consumers should also know when an image or video is authentic, and there are helpful industry standards being developed to address this concern. If a watermark has been affixed to an authentic image, it should not later be removed.

Question 4. What role should platforms play in detecting AI-generated content on their sites and in informing consumers when content is AI-generated content?

Answer. BSA represents enterprise software companies that provide business-to-business (B2B) products and services that help their business customers utilize AI.

In this B2B capacity, BSA member companies generally do not interact directly with individual consumers (unlike consumer-facing platforms). While the most suitable way to achieve the goals of transparency with consumers with respect to AI-generated content will vary depending on how a business uses AI and how it interacts with individual consumers, many consumer-facing companies may be able to leverage open standards that help consumers trace the origin of different forms of media, such as the standards being developed by CAI and C2PA.

RESPONSE TO WRITTEN QUESTION SUBMITTED BY HON. BEN RAY LUJÁN TO VICTORIA ESPINEL

Question. How can a sector-or use-case specific approach improve AI transparency?

Answer. AI is used in different ways by different sizes of companies that operate in different industries. AI policies—including policies that promote transparency—should reflect that these different types of uses of AI create different risks. A use-specific approach is helpful for promoting transparency because different types of information will be relevant to individual consumers depending on the type of AI system being used and the purpose of the system.

Regulations should account for these different uses, focusing on those that pose the highest risk to individuals—those that determine eligibility for housing, education, employment, credit, healthcare, or insurance. It is important to note that many uses of AI, such as optimizing manufacturing or supply chain logistics, will not be consumer-facing and, therefore, applying consumer-focused transparency requirements may not be appropriate.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. JOHN HICKENLOOPER TO VICTORIA ESPINEL

International AI Regulation:

Question 1. The international community is closely watching what guardrails the United States establishes on high-risk AI systems. At the same time, the international community is moving forward with their own versions of legislation or frameworks to promote responsible use and development of AI systems. For example:

- The European Union is moving forward on passing their AI Act
- The United Kingdom is hosting an inaugural AI Summit this November focused on AI safety
- The United Nations and OECD are also moving forward with AI safety principles

Could you discuss the importance of the U.S. demonstrating international leadership on AI and building international consensus on our principles with our international allies?

Answer. The United States has a unique opportunity to lead the international conversation on the responsible use and development of AI because it is the home of leading AI innovators, many of which are BSA members. The United States should foster an environment that enables AI innovation to continue to thrive while also establishing protections that address its risks. The United States has recognized the enormous benefits of working with other countries on AI governance issues and is already participating in a range of global efforts including through the G7 Hiroshima AI project and the Organisation for Economic Co-Operation and Development (OECD). The U.S. voice in those efforts will be stronger if the United States adopts national AI legislation that creates clear guardrails for how companies develop and deploy high-risk AI systems. Other jurisdictions, such as the EU, are moving forward with their approaches. Now is the time for Congress to act.

Federal Capabilities:

Question 2. Several Federal agencies and offices have launched initiatives to advance safe and trustworthy AI, including the Department of Commerce, Department of Energy, and the Federal Trade Commission. In your view, which Federal agencies are equipped to implement aspects of AI policy? What strengths do these agencies offer to support trustworthy AI development or deployment, respectively?

Answer. Because companies in all industries are adopting AI, the use of AI technologies will implicate areas covered by a range of agencies, *e.g.,* the Food and Drug Administration, the Federal Trade Commission, the Department of Justice, the Consumer Financial Protection Bureau, the Department of Commerce, and the National Highway Traffic Safety Administration. Each agency has a role to play in providing domain-specific expertise on issues falling within its jurisdiction. It is also important to ensure that Federal agencies coordinate with each other in their approach to AI, to promote consistency in the Federal government's overall AI policy. For example, the National Institute of Standards and Technology (NIST) can play a role in supporting risk management efforts that leverage its work in developing the AI Risk Management Framework (NIST AI RMF).

Consumer Notice:

Question 3. Generative AI systems are becoming more prevalent in the form of customer support chatbots and personal assistants. Should consumers be informed whether they are interacting with a generative AI system? What is the most effective way to plainly and easily notify consumers of this interaction?

Answer. Transparency is an important part of building trust in AI systems. One way to promote transparency is for companies that use customer support chatbots and personal assistants to inform consumers that they are interacting with an AI chatbot. In practical terms, an AI chatbot could inform the consumer that it is powered by AI in response to the first query a consumer submits.

Review Boards:

Question 4. Some companies have established Internal Boards to assess and determine whether an AI system is developed with safety, risk mitigation, and ethics in mind. The majority of Internal Boards consist of personnel within the company, and may not always include independent experts. In general, what is the typical review process for an AI system conducted by engineers or an Internal Board within a member company? If engineers or Internal Board members draw different conclusions after an AI system review, how are disputes resolved among these parties?

Answer. Companies adopt varied approaches to AI governance, including the adoption of Internal Boards or other bodies that oversee the development and use of AI-related products and services. For example, companies may establish a process for reviewing AI systems that entails an initial review by designated personnel, and then a subsequent and more detailed review for systems identified as having a higher risk profile. That subsequent review may be undertaken by an internal review board or other body charged with oversight of the company's development and use of AI-related products and services. That body is typically empowered to identify risk mitigations for engineers to implement. BSA has promoted the use of risk management programs that build on the NIST AI RMF. For example, companies can apply the NIST AI RMF's "govern" function to establish policies, processes, procedures, and practices across their organization to map, measure, and manage AI risks.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. PETER WELCH TO VICTORIA ESPINEL

Question 1. The European Union's (EU) Artificial Intelligence (AI) Act is set to become the world's first comprehensive legal framework for AI. Although EU regulations in AI and privacy only apply within the EU market, U.S. Internet companies that operate internationally tend to apply EU regulations everywhere they operate, which is known as the "Brussels Effect."

a. As Congress considers comprehensive AI legislation, what lessons can be learned from the enactment and implementation of the EU's AI Act, particularly in the realm of consumer protection and transparency?

Answer. Trilogue negotiations on the EU AI Act are ongoing, so the text of the legislation is not final. But one important aspect of the EU AI Act is its risk-based approach to AI regulation. The EU AI Act focuses on uses of AI that create specific concerns and attempts to address those concerns from a risk management perspective. This approach ensures that regulations focus on high-risk uses of AI, such as

credit scoring, and does not impose overly burdensome requirements on uses of AI that create low risks to individuals. We have encouraged the EU to maintain this risk-based approach, which we believe is important to any AI regulation.

The transparency measures in the EU Council's version of the EU AI Act are also instructive. The EU AI Act ensures that developers of high-risk AI systems provide deployers with information regarding the capabilities and limitations of the system, including any known or foreseeable circumstances that may lead to risks to health, safety, or fundamental rights, and human oversight measures. This ensures that key information about an AI system is provided by companies developing the AI system to the companies using that system.

In addition, the EU AI Act requires developers to design AI systems to inform individuals that they are interacting with AI systems. In addition, users of systems that generate deepfakes must disclose that the content has been artificially generated, with exceptions for criminal law enforcement and artistic content. We recognize the importance of these issues and generally agree that consumers should be told when they are interacting with AI chatbots, and support efforts to help consumers know when an image is authentic.

b. What impact do you expect EU's AI policy to have on technological innovation in the U.S.?

Answer. It is not yet clear what impact the EU's AI policy will have on technological innovation in the United States because the text of the EU AI Act is still being finalized. The original proposal would establish a risk-based approach to regulating AI, and BSA believes that a risk-based approach is the right way for policymakers to address AI. Focusing obligations on high-risk uses of AI ensures that low-risk uses of AI are not subject to overly burdensome requirements that may impede the ability of companies to continue to innovate, while creating important guardrails on high-risk uses of AI technologies. The full impact of the legislation will also be determined by how other outstanding issues are resolved, including how responsibilities are allocated along the AI value chain.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. TED CRUZ TO VICTORIA ESPINEL

Question 1. ***Impact Assessments.*** In your testimony, you stated that AI legislation should require companies to "conduct annual impact assessments for high-risk uses of AI."

You note that impact assessments are "already widely used in a range of other fields." However, the frequency of their use does not equate to good policy. For example, when I hear "impact assessment," I am reminded of the National Environmental Policy Act (NEPA)—a policy that has received criticism from both sides of the aisle for imposing crushing requirements on American businesses. While initially concise, NEPA assessments have ballooned to over 600 pages on average, with appendices alone averaging over 1,000 pages.[1] As you can imagine, the time investment to produce a document like this is significant, averaging 4.5 years and, in some instances, stretching beyond 17 years.[2] This protracted and resource-intensive process often derails or delays vital public projects and increases costs substantially.[3]

Given how rapidly AI is advancing and how revolutionary it promises to be, we do not have years to waste on bureaucratic processes that are ineffective, subjective, and hinder the adoption of important innovation.

a. Recognizing that many of the ideas for AI impact assessments mirror NEPA requirements, how would you ensure that they are less expensive and time-consuming than NEPA? Please provide concrete evidence of the individual impact of mandatory impact assessments on large, medium, and small businesses.

Answer. Impact assessments are an important accountability tool and should be designed to focus on a key set of issues. Clearly focusing an impact assessment on a core set of information can reduce the burden on companies obligated to perform them. For example, an impact assessment for developers of an AI system would generally focus on the intended use of the AI system, the known limitations of the system, the known likely and specific high risks that could occur and the steps taken

---

[1] Council on Environmental Quality, "Length of Environmental Impact Statements (2013–2017)," Executive Office of the President, July 22, 2019. *https://ceq.doe.gov/docs/nepa-practice/CEQ_EIS_Length_Report_2019-7-22.pdf.*

[2] Eli Dourado, "Why are we so slow today?," The Center for Growth and Opportunity, Mar. 12, 2020. *https://www.thecgo.org/benchmark/why-are-we-so-slow-today.*

[3] Ibid.

to mitigate those risks, an overview of the type of data used to train the system, and a summary of how the system was evaluated prior to sale. Legislation can create a clear and focused set of requirements for impact assessments, which will help maximize their use as an accountability tool.

Notably, impact assessments may also be leveraged across multiple jurisdictions. For example, ten states have privacy laws that will require privacy impact assessments for certain activities. In many cases, a company operating across state lines will be able to invest its resources in creating a process for conducting impact assessments that can comply with requirements across these jurisdictions.

In NEPA, mitigating potential negative impacts often involves changing the way a particular action is undertaken to minimize environmental harm. In AI, mitigation could involve changing an algorithm, employing differential privacy techniques, or even reconsidering the deployment of a particular AI system.

b. Please list all companies you represent that support impact assessments being required by the government. Do these companies support conducting them "annually"?

Answer. BSA supports legislation that requires annual impact assessments for companies that develop or deploy high-risk AI. A list of BSA member companies is available at bsa.org/membership.

c. Do you view internally-conducted, voluntary impact assessments as valuable? Why or why not?

Answer. Yes. Impact assessments are conducted by internal corporate teams and are important accountability tools because they are instrumental in helping companies identify and mitigate risks associated with an AI system. Many state privacy laws, including the recently-enacted privacy law in Texas, require companies to undertake internally-conducted impact assessments when they engage in certain activities. Impact assessments are designed to identify risks associated with those activities and to drive changes by the internal product teams responsible for developing a product or service, so that those teams can mitigate the identified risks. When impact assessments are conducted internally, they can also take into account information that may be difficult to disclose to third parties without raising trade secret and privacy concerns.

Although voluntary impact assessments enhance the accountability of the companies that perform them, BSA supports legislation that would require impact assessments for all companies that develop or deploy high-risk AI, which we believe will help build trust and confidence in adoption of AI technologies.

Question 2. ***Regulation.*** In your testimony, you said that "Congress should not wait to enact legislation that creates new obligations for companies that develop and use AI in high-risk ways." As I noted in my opening statement submitted for the record, the U.S. already has laws and regulations that protect consumers from harms, such as the Civil Rights Act of 1964 and the Fair Credit Reporting Act. The use of any AI system would still be subject to these laws.

a. Given the many existing regulations that we can enforce for AI systems across all sectors, why do you think the time is now to create new regulations?

Answer. We recognize that existing laws apply to AI technologies, and BSA has encouraged agencies to undertake a review of existing laws to identify potential gaps in those authorities. At the same time, BSA supports legislation that would require companies that develop or deploy high-risk AI systems to conduct impact assessments. Those assessments will help companies identify and mitigate risks that arise from an AI system, including the potential that the system may violate an existing law or may have a significant impact on a consumer's privacy. The United States should have a role in how AI rules are developed globally, and the window to help shape those conversations is closing.

b. Was this your opinion prior to the generative AI boom with ChatGPT at the end of last year, and if not, can you please describe how it has changed?

Answer. In 2021, BSA began calling for legislation that requires companies that develop or deploy high-risk AI systems to conduct impact assessments. We continue to believe that this type of legislation is critical to identifying and mitigating risks of AI systems and ultimately building trust in the technology.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. AMY KLOBUCHAR TO DR. RAMAYYA KRISHNAN

Artificial Intelligence (AI) and Elections: Artificial intelligence (AI) has the potential to completely upend campaigns and elections unless there are rules of the road in place. I introduced bipartisan legislation with Senators Hawley, Coons, and Col-

lins to prohibit the use of deceptive AI-generated audio, images, or video in political ads because some of this content needs to be banned altogether–and I also introduced the *REAL Political Ads Act* with Senators Booker and Bennet, which is led in the House by Rep. Yvette Clarke (D–NY), to require disclaimers on political ads containing AI-generated images or video so voters are aware when this technology is used in our elections.

Question 1. Do you agree that without giving people information to determine whether an image or video was created by AI, that generative AI poses a risk to our free and fair elections?

Answer. We need to give citizens information that they can use to determine whether an image or video was produced synthetically via AI. So content creators need to label content and provide detection tools that can process such labels to alert citizens about synthetic media use in images and video. Note that the proposal is not taking a position on the content that was created but rather informing the citizen about content provenance (how it was created and where it came from).

Question 2. As AI makes spreading deceptive content cheaper and easier, what steps do you think should be taken to protect our democracy?

Answer. First, we need to protect our election systems by getting stakeholders to agree to a code of conduct and charge and resource a lead agency (*e.g.,* Federal election commission) to enforce this code of conduct. Second, we need to educate citizens (the Block Center for Technology and Society (*https://www.cmu.edu/block-center/*) at Carnegie Mellon has created a voters guide to help educate citizens. Please see *https://www.cmu.edu/block-center/responsible-ai/genai-voter-guide.html*) and harness both Federal government department resources as well as resources from academia and civil society to assist with public digital literacy education and disinformation awareness in the run up to the 2024 U.S. elections. Third, we need content labeling standards and detection tools. Example of a content labeling standard is synthid from Google. The C2PA (*https://c2pa.org*) alliance is promoting a content provenance standard. However, there is no existing adopted standard for closed and open source models as well as for devices (cameras and video units) that create the content. Having a content labeling standard will be a major step forward.

We also need to monitor and use technology when content labels are not available. The state of the art of deep fake identification in the absence of labels is constantly improving. The following papers provide a state of the art review:

https://www.nature.com/articles/s41598-023-34629-3

https://arxiv.org/ftp/arxiv/papers/2103/2103.00484.pdf

https://par.nsf.gov/biblio/10356295

https://www.pnas.org/doi/full/10.1073/pnas.2110013119

Scams, Fraud, and Phishing: According to experts it only takes three seconds of audio to clone a voice. That means with just a small clip of audio taken from social media or another source, scammers can produce an almost perfect voice match using easily accessible artificial intelligence (AI) programs.

Question 3. What responsibilities do developers and deployers of AI systems have to prevent their technologies from being used to defraud consumers?

Answer. We need a robust synthetic media code of conduct. Hand in hand with this code of conduct is technology to establish content provenance. Audio watermarking technology can mitigate fraud or scams as long as content labeling and detection requirements are in place. The U.S. Government should invest in research that will continually improve digital content provenance and watermarking and as well in methods to detect synthetic media enabled fraud and scams in order to stay ahead of scammers. An enforceable code of conduct with lead agencies resourced and charged with enforcement (*e.g.,* Federal election commission, CPFB, FTC) combined with technical tools will provide the incentives and norms for all developers and deployers to create a more trusted digital system.

Impact of AI on the Workforce: As we look to the future, it is critical to be able to measure the impact artificial intelligence (AI) will have on wages and economic opportunity. We have heard both positive and negative predictions about the impact AI will have on our economy. I am concerned we do not have enough data to inform effective policymaking.

Question 4. What types of data about the economic impact of AI do you think is important to collect to inform policy making?

Answer. The statistical agencies (BLS, BEA, Census) are great assets. They are comprehensive in that the data they collect span rural and urban geography. The data are also representative of the U.S. population. However, the agencies lack high

frequency data. Fast moving phenomena and disruptions require policy makers to have situational awareness.

Commercial data acquirers and providers (linkedin, workday and lightcast to name a few) have high frequency data but the data is neither comprehensive nor representative. A 2020 national academies report on the consumer food distribution system made very similar observations (see *https://www.ncbi.nlm.nih.gov/books/NBK561996/*) and called for the appropriate combination of government and private data sources to support specific policy maker use cases, An upcoming National Academies Report co-chaired by Block Center of Technology and Society leader Prof. Tom Mitchell (see *https://www.nationalacademies.org/our-work/automation-and-the-us-workforce-an-update*) will address the question you have posed in detail. In addition to supporting the recommendations in these reports, we need continued investments in research into how to make creative and innovative use of our increasingly digital world to measure economic and societal indicators relevant to policy making.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. BRIAN SCHATZ TO DR. RAMAYYA KRISHNAN

Question 1. Do people have the right to know when they are viewing AI-generated content or when they are interacting with an AI-chatbot?

Answer. Yes, I believe they do need to be informed that they are interacting with synthetic media generated by an AI. I called for this during my testimony to the committee.

Question 2. What are the dangers if individuals do not have a clear understanding that they are viewing or interacting with AI-generated content?

Answer. Our democracy is based on having access to trusted sources of information and beliefs founded on attributing what one sees or hears to the individual making these remarks. If a citizen is deliberately misled to believe something about a candidate, this will further erode trust in our election system. While I have illustrated the problem in the content of elections, commerce could be affected as well. Example: purchasing digital content that sounds like it was created or produced by a reputed star (*e.g.,* songs like Drake—please see *https://tinyurl.com/45aba3uu*).

Question 3. Do people have the right to know what model created the AI-generated content they are viewing or interacting with? Should generative AI systems affix information to AI-generated content reflecting that the content is generated by AI? Should that information be permanently affixed to the extent doing so is technically feasible?

Answer. We need a content labeling and detection system. I believe this should go hand in hand with a synthetic media code of conduct which is enforceable. Helping a citizen to be reliably informed that the content she is viewing (images/video) is synthetic or not will be a big step forward. Further capacity to inspect the content provenance (*e.g.,* which model/device created it and its modification history) would be relevant for auditors. The *https://c2pa.org/* is a industry-led coalition seeking to promote content provenance standards. Their solution calls for this information to be affixed to digital content.

Question 4. What role should platforms play in detecting AI-generated content on their sites and in informing consumers when content is AI-generated content?

Answer. If content standards existed, it would be straightforward for platforms to use the detection tool offered by the content labeling technology to inform consumers on the platform about whether the content they are interacting with is synthetic media generated by an AI. In the absence of such a standard, classifying content as AI generated or not is not perfect and errors of omission and commission can be committed. Given that detection is not perfect, reliance on some combination of technology tools and crowdsourced complaints, and use of oversight boards will likely be the response of platforms. They do not want or should not be seen as censoring speech. For example, platforms like Meta have oversight boards who provide advice to platforms such as in the case of doctored video of Pres. Biden. In that case, Meta stated ". . . it will remove a video if it "has been edited or synthesized . . . in ways that are not apparent to an average person, and would likely mislead an average person to believe a subject of the video said words that they did not say."

Question 5. The *AI Labeling Act,* which I introduced with Senator Kennedy, would require generative AI systems to add visible labels and metadata information identifying content as AI-generated. Please explain the importance of both consumer-facing labels as well as embedded labels and associated detection tools. Should these measures be required for all AI-generated content?

Answer. Requiring content labeling and detection tools will be an important step forward. I think this needs to go hand in hand with a synthetic media code of conduct which is enforceable by a lead agency with the resources and powers to do so (*e.g.,* FEC in the case of elections). There are technical solutions such as synthid from Google and the C2PA (*https://c2pa.org*) alliance that is promoting a content provenance standard.

However, there is no existing standard for closed and open source models as well as for devices (cameras and video units) that create the content in the first place. While there will always be some open source alternatives that do not comply with these requirements/standards, having all the major open source and closed source players voluntarily adopt these requirements will be a major step forward.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. BEN RAY LUJÁN TO DR. RAMAYYA KRISHNAN

Question 1. The explosion of generative AI has serious implications for the integrity of our information ecosystems. Should Congress require technology companies to label any AI-generated content with disclosures and metadata that includes information on how the content was created?

Answer. Requiring content labeling and detection tools will be an important step forward. I think this needs to go hand in hand with a synthetic media code of conduct which is enforceable by a lead agency with the resources and powers to do so (*e.g.,* FEC in the case of elections). There are technical solutions such as synthid from Google and the C2PA (*https://c2pa.org*) alliance that is promoting a content provenance standard.

However, there is no existing standard for closed and open source models as well as for devices (cameras and video units) that create the content in the first place. While there will always be some open source alternatives that do not comply with these requirements/standards, having all the major open source and closed source players voluntarily adopt these requirements will be a major step forward.

Question 2. Transparency is critical not only to protecting consumers, but also protecting artists whose work may have been used—often without permission—to train AI models. How can standardized documentation help consumers and oversight bodies such as the U.S. Copyright Office understand potential risks from AI systems?

Answer. I have called for "nutrition labels" to document what data are used in training an AI model and what rights the model developer has to the data that is being used for this purpose. While licensing offers a path forward to both creators of content as well as to model developers, the first step is having model developers declare and document what data they used and the sources of these data. At present, model cards or system cards provide high level statements of data that are used but there is no standard template that provides the level of detail required A recent proposal for a foundation model transparency index is a useful device to monitor how well model developers are doing on these dimensions (see *https://arxiv.org/pdf/2310.12941.pdf*).

Question 3. How can a coordinated U.S. government approach to AI transparency ensure U.S. leadership in responsible AI?

Answer. I am attaching a detailed policy memo on accountable AI from the Block Center for Technology and Society (please see *https://www.cmu.edu/block-center/responsible-ai/cmu_blockcenter_rai-memo_final.pdf*) in response to your question.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. JOHN HICKENLOOPER TO DR. RAMAYYA KRISHNAN

Federal Capabilities:

Question 1. Several Federal agencies and offices have launched initiatives to advance safe and trustworthy AI, including the Department of Commerce, Department of Energy, and the Federal Trade Commission. In your view, which Federal agencies are equipped to implement aspects of AI policy? What strengths do these agencies offer to support trustworthy AI development or deployment, respectively?

Answer. The NIST is a center of excellence of AI risk management. They have led the development of the AI RIsk Management Framework which provides via a feature called profiles a way to set acceptable risk thresholds for use cases. The AI RMF has wide support in industry. So NIST (which is in Commerce) is well positioned in partnership with agencies such as the Department of Energy and Department of Defense to create measurement, test and evaluation approaches and mecha-

nisms to operationalize AI development and deployment. These agencies have deep, multi-disciplinary strength in AI technology, socio-technical systems, security, privacy, harm analysis, etc.). The FTC could monitor the competitiveness of the U.S. market for AI technology and application to ensure it remains vibrant and vital.

Consumer Notice:

Question 2. Generative AI systems are becoming more prevalent in the form of customer support chatbots and personal assistants. Should consumers be informed whether they are interacting with a generative AI system? What is the most effective way to plainly and easily notify consumers of this interaction?

Answer. We should inform consumers that they are interacting with a generative AI vs. a human in a customer support type of application. This follows from our values about not fielding deceptive systems. It is difficult to prescribe the best way to notify the user. It will depend on context and the specific approaches are areas of active research.

More generally, there is research literature on humans and ai in decision making contexts. With predictive AI, researchers have studied "algorithm aversion" (humans rejecting use of AI) and recent work has looked at how to minimize this type of aversion (please see *https://dl.acm.org/doi/fullHtml/10.1145/3544548.3581253*). In contrast to predictive AI, generative AI appears to be perceived differently (see *https://papers.ssrn.com/sol3/papers.cfm?abstract_*id=4453958).

Addressing Vulnerabilities:

Question 3. Many AI developers test their systems prior to release for flaws and vulnerabilities. One common method is "red-teaming", where experts manually look for problematic behavior that needs to be fixed. What are some limitations of "red-teaming" and how will this practice need to expand and evolve to ensure AI systems remain safe over time?

Answer. "Red teaming" in the context of AI systems is not a well defined term. Recent examples appear to focus on "prompt hacking" to produce bad behavior from the model as the sole focus of red teaming. This is too narrow. Borrowing from its roots in cybersecurity, what is required is AI testing done prior to deployment keeping in mind the needs of both AI safety (reducing harm caused by reliability errors, misuse or malfunction) and AI security (defending systems against malicious actors (*e.g.,* data poisoning attacks)). It is important that the scope of this effort include impact analysis of the use of the AI system (the system may include human and AI components) A memo on accountable AI from the CMU Block Center for Technoloigy and Society is available at *https://www.cmu.edu/block-center/responsible-ai/cmu_blockcenter_rai-memo_final.pdf*) and a report on red teaming from the Frontier Model Forum is a useful resource (*https://www.frontiermodelforum.org/uploads/2023/10/FMF-AI-Red-Teaming.pdf*). Post deployment there is need for an institutional infrastructure (pleasee the call for a trust infrastructure in my senate testimony at *https://www.commerce.senate.gov/services/files/96B6B41C-9335-43AF-9DB1-1231AF66C493*) much like there is with the computer emergency response team (CERT) for information security. We need a CERT for AI which will catalog ai failures, incidents and vulnerabilities and conduct forensic analysis of these to inform model developers who will develop model updates to address these failures to ensure safer model deployments. Creating safe AI ecosystems will require additional research and infrastructure to support AI safety and security through the AI lifecycle.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. PETER WELCH TO DR. RAMAYYA KRISHNAN

Question 1. Earlier this year, Senator Bennet and I reintroduced the Digital Platform Commission Act, first-ever legislation to create an expert Federal agency to provide comprehensive regulation of digital platforms to protect consumers, promote competition, and defend the public interest. We updated the text to incorporate the regulation of AI and social media, which includes the mandate, jurisdiction, and tools to develop and enforce rules for a sector that has gone virtually unregulated.

a. Similar to the government's strategy towards the regulation of social media and data privacy, the proposed framework for AI regulations incorporates a fragmented approach that includes multiple different Federal agencies. Do you believe that a fragmented approach is the most effective way to regulate AI?

Answer. There are two requirements. The first is to govern AI in specific use cases and we have agencies with oversight authority with existing laws (*e.g.,* EEOC, CPFB etc.). These agencies have to be resourced with expertise and AI capabilities

to do their job. The second requirement is to coordinate AI standards globally and serve as a single point of contact to work with the EU and other governments and their AI regulatory agencies. This requires a new agency/commission. This commission will need to have both technology and policy expertise that spans multiple use cases. So I think we need a federated approach that provides for AI expertise and capability in existing agencies with oversight over consequential use cases while standing up a new agency/commission to coordinate globally AI (tech) policy.

b. How could the Federal government and U.S. technology companies benefit from the creation of a single government commission focused on the regulation of AI and digital platforms, such as what has been proposed in my & Senator Bennet's Digital Platform Commission Act?

Answer. The principal advantage is the creation of a center of excellence in AI technology and AI policy with a deep understanding of the AI oversight needed in critical use cases (*e.g.,* in lending, hiring, health care etc.). Tech companies and the government will look to this agency/commission as the source of expertise and information on AI policy within the U.S. and Globally, It could also play an important role in harmonizing AI policy with our key trading partners.

As noted above, this will require a clear scoping of mission and responsibility as well as close partnerships (data sharing, appointments) between the new AI agency/commission and existing agencies and commissions that will apply existing law to govern AI in use cases.

RESPONSE TO WRITTEN QUESTION SUBMITTED BY HON. TED CRUZ TO
DR. RAMAYYA KRISHNAN

Question. ***Labelling.*** In your testimony, you said that "Congress should require closed-source and open-source models to actually create this watermarking label and a detection tool to go with this label."

a. For the companies that are already looking into watermarking and detection methods, how often do existing, state-of-the-art tools incorrectly identify authentic images as AI-generated?

Answer. My testimony called for model developers to adopt content labeling (*e.g.,* watermark) and to provide detection tools that will use these labels to inform a citizen that they were viewing ai generated content. To elaborate, this technology capability needs to go hand in hand with a synthetic media code of conduct which is enforceable by a lead agency with the resources and powers to do so (*e.g.,* FEC in the case of elections). Examples of content labeling standards is synthid from Google. The C2PA (*https://c2pa.org*) alliance is promoting a content provenance standard. However, there is no existing adopted standard for closed and open source models as well as for devices (cameras and video units) that create the content.. Having a content labeling standard will be a major step forward.

We also need to monitor and use technology when content labels are not available. The state of the art of *deep fake identification in the absence of labels* is constantly improving. The following papers provide a state of the art review:

https://www.nature.com/articles/s41598-023-34629-3
https://arxiv.org/ftp/arxiv/papers/2103/2103.00484.pdf
https://par.nsf.gov/biblio/10356295
https://www.pnas.org/doi/full/10.1073/pnas.2110013119

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. BRIAN SCHATZ TO
SAM GREGORY

Question 1. Do people have the right to know when they are viewing AI-generated content or when they are interacting with an AI-chatbot?

Answer. There are already a range of forthcoming and proposed laws, voluntary commitments, and high-level guidance in the United States and Europe which require foundation model developers, generative AI tool builders, distributors, and others to notify people when they are interacting with AI-generated content, including AI chatbots. Most recently, in September, Senators Blumenthal (D–CT) and Hawley (R–MO), announced a bipartisan framework for AI legislation intended to be a 'comprehensive legislative blueprint'.[1] Among other proposals, the framework

[1] Richard Blumenthal and Josh Hawley, Bipartisan Framework for U.S. AI Act, *https://www.blumenthal.senate.gov/imo/media/doc/09072023bipartisanaiframework.pdf*

states that 'users should have a right to an affirmative notice that they are interacting with an A.I. model or system.'[2] In July, the AI Labeling Act introduced in the U.S. Senate by Senators Schatz (D–HI) and Kennedy (R–LA) mandates visible labels for image, video, audio, multimedia and text.[3] Also in July, seven leading AI companies agreed to the U.S. White House voluntary commitments, which are aimed at moving toward the safe, secure, and transparent development of AI technology, including committing to earning people's trust by disclosing when content is AI-generated.[4] These voluntary commitments include developing and deploying mechanisms that enable people to understand if audio or visual content is AI-generated, including robust provenance, watermarking, or both, for AI-generated audio or visual content. Earlier this year, the AI Disclosure Act, introduced in May by Representative Torres (D–NY), also required generative AI output to include a disclaimer.[5]

Signatories to the European Union's Code of Practice on Disinformation[6], another voluntary framework, include companies such as Abode, Google, Microsoft, and TikTok, among others. Companies that have signed on to the voluntary EU Code of Practice on Disinformation have agreed to a similar commitment, and the EU's Commissioner Věra Jourová has recently called on these companies to label AI-generated content.[7] The European Union AI Act,[8] which is expected to come into force in 2024, also includes provisions that place transparency requirements on AI systems. The AI Act is likely to require AI systems that are used to generate or manipulate images, audio, or video content that resembles authentic content to include disclosures that notify people that the content was generated through automated means.

All these measures reflect a necessity consistently identified in WITNESS' consultations and research:[9] it is not possible to place the burden on consumers and citizens to 'spot' that they are interacting with an AI-chatbot, or viewing or listening or otherwise consuming AI-generated or edited content. People do have a right to know they are viewing or otherwise consuming (for example, listening to) substantively AI-based content, particularly when this occurs in formats, contexts or settings where a reasonable person would not expect AI to be used. Similarly they should have a right to know they are engaging with an AI-chatbot, and to be regularly reminded of this.

However, any such measures need to be implemented recognizing the complexities of explaining how AI is employed when its usage may be partial or an accustomed and normalized use–for instance when dyslexic people or non-native speakers of a language use AI to help make a piece of text clearer, or when a photograph is edited with AI. Similarly, the format in which this information is disclosed to the consumer should also be accessible and interpretable for multiple audiences, acknowledging the borderless nature of content dissemination and the needs of a diverse consumer base. Proposals for a direct disclosure should not require the creators of AI-based content to divulge personally identifiable data as part of the 'provenance' of AI content. AI systems should never capture by default personally identifiable information nor include other metadata that may be able to identify the individual operating the artificial intelligence system and generating the outputs, unless the user has been notified and they have opted-in to include this information.

Question 2. What are the dangers if individuals do not have a clear understanding that they are viewing or interacting with AI-generated content?

---

[2] ibid.

[3] AI Labeling Act of 2023, S. 2691, 118th Cong, 1st Session, (2023) *https://www.congress.gov/118/bills/s2691/BILLS-118s2691is.pdf*

[4] The White House, Fact Sheet, *Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI,* 21 July 2023, *https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/*

[5] AI Disclosure Act of 2023, HR. 3831, 118th Cong, 1st Session, (2023) *https://www.congress.gov/118/bills/hr3831/BILLS-118hr3831ih.pdf*

[6] European Commission, *2022 Strengthened Code of Practice on Disinformation,* 16 June 2022 *https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation*

[7] Foo Yun Chee, *AI-generated content should be labelled, EU Commissioner Jourova says.* Reuters, June 2023, *https://www.reuters.com/technology/ai-generated-content-should-be-labelled-eu-commissioner-jourova-says-2023-06-05/*

[8] European Parliament, *European Artificial Intelligence Act, https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html*

[9] For an overview of this research see: *www.gen-ai.witness.org*

Answer. As noted in my written testimony,[10] there is a rapid increase in the realism of content produced by generative AI. In the case of AI systems that produce realistic text, this is of particular importance given the potential privacy implications. For instance, users may inadvertently reveal sensitive information. Even if that data is not sold to a third party, the information may still be saved by the system for improving performance or further research. Additionally, these systems tend to 'hallucinate', providing inaccurate answers with an aura of credibility. More broadly, consumers may be deceived by audiovisual AI content that seems realistic, and fall prey to fraud and scams. As a result, we are seeing how the increasing volume of synthetic content is already undermining consumers' confidence in human-generated content.

Threats from inadequate signaling when content has been generated via AI are disproportionately experienced by those already at risk. In WITNESS's research we see how threats from synthetic media impact those already at risk because of their gender, sexual orientation, profession, ethnicity or social group identity. For example, for years now women have been a primary target by those misusing deepfake technologies to create non-consensual sexual imagery. In our global consultations, different stakeholders have also pointed to the threat of the use of synthetic media to attack and undermine civil society and independent media, and how it is misused to dismiss authentic information with claims it is falsified. This is so-called plausible deniability or the "liar's dividend".[11] This pressure places a strain on already under-resourced local and national newsrooms and community leaders responsible for verifying digital content. With hyperbolic rhetoric as well as the realities of advances in generative AI undermining trust in content we encounter, human rights defenders, journalists and civil society actors will be among the most impacted by generative AI. My written statement has further examples of accidental, misuse, supply chain and structural harms.

Question 3. Do people have the right to know what model created the AI-generated content they are viewing or interacting with? Should generative AI systems affix information to AI-generated content reflecting that the content is generated by AI? Should that information be permanently affixed to the extent doing so is technically feasible?

Answer. In my written statement and in my answer to Question 3 (Senator Hickenlooper) I explain the different technical approaches to provide consumers with information about the content they are viewing, and the limitations of invisible watermarking, cryptographic signatures and labeling.

In terms of visible disclosure, synthetic content co-exists with non-AI generated content, and material that is a blend of the two (for instance a video originally shot in a phone with synthetic audio in only part of it, or a photograph with elements in painted or out painted). For this reason, simply indicating the presence of AI may not be that helpful for consumers without a more nuanced explanation of how and where AI has been used. Additionally, most visible labels are easily removable or non-maliciously withdrawn–for example a user may cut a clip from a longer video made with AI that has a disclaimer at the start, if a third user takes that video for further editing or circulating, the disclaimer will not be attached to it. This also raises questions about the accessibility and interpretability of different formats by different audiences in a global information ecosystem.

Concerning other solutions that affix information such as metadata, these approaches should never capture by default personally identifiable information nor include other sensitive data that may be able to identify the user of an AI generator tool–unless they have opted-in to include this information. Permanently affixing provenance information when synthetic and non-synthetic artifacts co-exist in a piece of media brings up more complexity than in situations when the content is binary synthetic or not. Lessons from platform policies around the use of 'real names'[12] tell us that many people—for example, survivors of domestic violence—

---

[10] Sam Gregory, *The Need for Transparency in Artificial Intelligence,* Testimony Before the U.S. Senate Committee on Commerce, Science and Transportation, Subcommittee on Consumer Protection, Product Safety and Data Security, September 2023, *https://www.gen-ai.witness.org/wp-content/uploads/2023/09/Senate-Commerce-written-testimony_FINAL.pdf*

[11] Robert Chesney and Danielle Keats Citron, Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security, 107 California Law Review 1753, July 2018, *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3213954*

[12] Jillian York and Dia Kayyali, *Facebook's 'Real Name' Policy Can Cause Real-World Harm for the LGBTQ Community,* Electronic Frontier Foundation, 16 September 2014, *https://www.eff.org/deeplinks/2014/09/facebooks-real-name-policy-can-cause-real-world-harm-lgbtq-community*

have anonymity and redaction needs that can cause life threatening situations if not respected.

Question 4. What role should platforms play in detecting AI-generated content on their sites and in informing consumers when content is AI-generated content?

Answer. Platforms play an important role in detecting and disclosing AI-generated content to people using their sites. People and companies working across the AI pipeline all have a duty to insert safeguards and address the harms their work can bring. It is a recipe for failure if the responsibility is left solely on end-users to determine if the audiovisual content they are viewing is AI-generated, as well as the larger context of the content they are consuming.

However, it should be noted that platforms are not able to implement after-the-fact detection at scale in a reliable manner. Generative AI detection for both audiovisual and textual content is also subject to adversarial attacks, and it is not realistic to set an expectation that puts the burden solely on distribution platforms to detect AI-generated content. For detection to have a real impact, it would require collaboration and standard setting across the AI pipeline, including model developers. Currently there is not a generally used interoperable standard for understanding how AI was used in content creation, and for indicating this to individuals. A number of company specific watermarking initiatives have been started, providing indications of AI-generation and point-of-creation. Similarly, the Coalition for Content Provenance and Authenticity (C2PA) has introduced an open standard for metadata-based provenance indicating how a piece of content is created, edited and shared. However, there is as yet no widely used interoperable standard for either watermarking or metadata-based provenance.

As highlighted in my answer to Question 1 (Senator Schatz), there are already a range of forthcoming laws, voluntary commitments, and high-level guidance in the United States and Europe which require distributors such as social media platforms to disclose AI-generated content on their sites. Any legislative proposals should ensure to promote standardization of watermarking as well as other disclosure techniques such as metadata-based provenance that facilitate better detection and allow for interoperability across platforms.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. BEN RAY LUJÁN TO SAM GREGORY

Question 1. To what extent are deepfakes and AI-generated disinformation already being created in Spanish and other non-English languages?

Answer. AI-generated or manipulated deepfakes and disinformation are already prevalent in Spanish and other non-English languages, and we believe that the number of cases and their sophistication will continue to escalate in the short term, within and outside of the United States.

In March, WITNESS started piloting a Deepfake Rapid Response Force (the 'Force'),[13] connecting local journalists and fact-checkers across the globe with leading AI-media forensics experts. The Force can take cases of suspected audiovisual synthetic content in different languages, and it has recently covered cases in Spanish, Arabic and Russian. It is notable that a number of the cases that have been received at WITNESS and escalated to the Deepfake Rapid Response Taskforce have been of audio content for which we have a reasonable suspicion to believe that it has been AI generated or manipulated. Generative AI and deepfake usage has also been recently seen in other non-English speaking political contexts, such as Slovakia,[14] Brazil[15] and Ukraine,[16] and reflects the ease of production of audio-realistic AI-generated content.

In our recent consultation with stakeholders in Colombia, participants pointed out that they were seeing an increasing number of deepfakes and AI-generated content, especially in the context of elections. Some of the examples highlighted were the

---

[13] For an example of this work, see: *https://restofworld.org/2023/indian-politician-leaked-audio-ai-deepfake/*

[14] Morgan Meaker, *Slovakia's Election Deepfakes Show AI is a Danger to Democracy,* Wired, 3 October 2023, *https://www.wired.com/story/slovakias-election-deepfakes-show-ai-is-a-danger-to-democracy/*

[15] *Gabriela Coelho* and *Gabriel Hirabahasi, Cármen Lúcia manda remover vídeos contra Lula com "deep fake",* CNN Brasil, 28 October 2022, *https://www.cnnbrasil.com.br/politica/carmen-lucia-manda-remover-vid eos-contra-lula-com-deep-fake/*

[16] Bobby Allyn, *Deepfake video of Zelenskyy could be 'tip of the iceberg' in info war, experts warn,* NPR, 16 March 2022, *https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia*

fake news avatars lauding an alleged tourism boost in Venezuela,[17] or the deepfake in favor of Mexican presidential candidate, Xóchitl Gálvez.[18]

Question 2. Are technology companies investing enough resources into making sure AI systems work equally well in non-English languages?

Answer. In WITNESS experimentation and consultations we have corroborated what different studies [19] have already shown: most generative AI systems work best in English, they work well in high-resourced languages such as Spanish, and they underperform in low-resourced languages such as Punjabi [20] or Tamil.[21] For AI systems to work well in non-English languages, companies need to invest in balancing supply chain inequities, such as the fact that there is simply more scrapable data to train models in English than in other languages. Companies should also employ or work with speakers of other languages to design and fine-tune models. At a broader scale, the use of large language models and AI systems in content moderation is rarely sufficiently localized to contexts outside English-speaking U.S. and Europe [22], nor is it reinforced with well-resourced, paid and supported human content moderation in non-English languages–with minority populations and vulnerable communities bearing the impact.

Question 3. How are non-English speaking communities uniquely vulnerable to mis- and disinformation generated and spread artificial intelligence?

Answer. As remarked in my answer to Question 2 (Senator Schatz), threats from synthetic media and generative AI used deceptively and maliciously will likely disproportionately impact those who are already at risk because of their gender or gender identity, sexual orientation, profession, ethnicity, belonging to a social group or because of their language. This vulnerability reflects existing patterns of misinformation, disinformation and harassment. For instance, generative AI tools make it easier to create content that promotes hatred towards minorities, and to produce it at scale. Additionally, as mentioned in my answer to Question 2 (Senator Luján), generative AI underperforms with non-English languages in comparison to English, and there is an investment gap in both AI-enabled and human content moderation that is not conducted in the English language.

More importantly, it is worth noting that the general mistrust of online content may make it easier to undermine or dismiss as 'fake' digital content from minorities like non-English speaking communities in the US, putting a higher burden on them to prove that their content is real or trustworthy.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. JOHN HICKENLOOPER TO SAM GREGORY

Open-Source AI Models:

Question 1. Open-source AI models could benefit consumers by promoting transparency and enabling audits for accountability. They could also create risks, like in a recent case where image watermarks were removed by modifying only one line of open-source code. Mr. Gregory, please explain both the benefits and risks of open-source AI models. How do open-source foundation models benefit consumers and the general public? In what ways could fine-tuned open-source models potentially harm consumers, and how could we proactively mitigate these risks?

Answer. Many of the communities and contexts in which I work have experienced in the past, and are still experiencing, the impact of technological concentration of power. They are currently observing the patterns of AI tools developed in the Global North that produce stereotypical images of their populations, underperform in their

---

[17] *Joe Daniels* in Bogotá and *Madhumita Murgia,* Deepfake 'news' videos ramp up misinformation in Venezuela, Financial Times 17 March 2023, *https://www.ft.com/content/3a2b3d54-0954-443e-adef-073a4831cdbd*

[18] Rodrigo Soriano, *Blanqueada y rejuvenecida: los videos de Xóchitl Gálvez creados con inteligencia artificial desatan la polémica en las redes,* El Pais, 25 July 2023, *https://elpais.com/mexico/2023-07-25/blanqueada-y-rejuvenecida-los-videos-de-xochitl-galvez-creados-con-inteligencia-artificial-desatan-la-polemica-en-las-redes.html*

[19] Viet Dac Lai *et al.,* Chatgpt beyond english: Towards a comprehensive evaluation of large language models in multilingual learning, 2023, arXiv preprint, *https://arxiv.org/abs/2304.05613*

[20] Alison Snyder, AI's language gap, Axios, 8 September 2023, *https://www.axios.com/2023/09/08/ai-language-gap-chatgpt*

[21] Andrew Deck, *We tested ChatGPT in Bengali, Kurdish, and Tamil. It failed,* Rest of World, 6 September 2023, *https://restofworld.org/2023/chatgpt-problems-global-language-testing/*

[22] Gabriel Nicholas and Aliya Bhatia, *Lost in Translation: Large Language Models in Non-English Content Analysis,* Center for Democracy and Technology, 23 May 2023, *https://cdt.org/insights/lost-in-translation-large-language-models-in-non-english-content-analysis/*

languages, and do not have an appropriate understanding of content or context for content moderation. I have also highlighted in my answer to Question 2 (Senator Luján) how supply chain inequities affect disproportionately minority communities and other groups that have been historically marginalized. From this perspective, open source models can help mitigate these biases as they allow for fine-tuning or further examination–however, it is worth noting that there are different levels of openness in a model.[23] Hence, an early ability to assess these technologies for bias is key. In this regard, open-source models can strengthen accountability, as they enable external researchers to more easily investigate, uncover and challenge model biases and threats. Broadening access to artifacts such as models and training datasets also allows researchers and users to better understand systems, conduct audits, mitigate potential risks, and find high value applications. For example, research about watermarking by the University of Maryland was conducted using OPT, an open-source language model developed and released by Meta. According to the scientists, the availability of the materials helped them understand the techniques' robustness. This scenario can in turn encourage other researchers to test and improve safety techniques.[24]

While open-source generative AI enables the customization of models to provide content and responses that are better suited to a range of local contexts, they also enable users to evade content moderation. Content controls are already inadequate for use at global scale, as mentioned above, but completely open systems make vulnerable people even more at risk to systems that have been fine-tuned to for instance, generate non-consensual sexual images, children sexual abuse material, biased outputs, and mis-and disinformation. We must not underplay these potential harms. The increasing volume, ease and access to generative AI and the possibilities for personalization that it brings, are likely to magnify this problem. Similarly, if models to detect synthetic content are open, they may also be rapidly subjected to adversarial testing, triggering a cat and mouse game of generation and detection that may make them less valuable.

To mitigate the risk of open models, Congress should consider encouraging scientific and research communities that are conducting active and robust testing. Gradient releases are particularly important in relation to detection tools, for instance, but there are other applications for which Congress should balance the benefits of openness. In sum, a multilayered technical, regulatory and societal approach will be necessary–one that involves and has the commitment of the different actors across the synthetic media pipeline, from model developers to deployers, to content distributors and platforms.

Since watermarking and other forms of durable disclosure that material is AI-generated may be removed easily from open-source code, legislators could consider how these watermark approaches could be implemented (while retaining critical protections for PII) at earlier stages of the AI pipeline, for example in training data or underlying models. They could also consider how both novel and existing legislation and regulation could be used to require the retention of watermarking even in open source.

Consumer Notice:

Question 2. Generative AI systems are becoming more prevalent in the form of customer support chatbots and personal assistants. Should consumers be informed whether they are interacting with a generative AI system? What is the most effective way to plainly and easily notify consumers of this interaction?

Answer. A general principle of informed disclosure is critical across generative-AI based content and interactions. Consumers should not be assumed to be aware that they are interacting with realistic or human-seeming AI-based content, chat-bots and systems. An appropriate approach to disclosure would involve notification, visibly or audibly, to the users when they begin interacting with a system, as well as durable signals embedded in the interaction that indicate origins in an AI-system. To-date it should be noted that there is not a standardized, robust method to provide detectable traces in text-generated content created by a chatbot. Congress could also consider if there are notable limitations to models that promote an anthropomorphic connection to the user such as adopting a particular persona (as

---

[23] Irene Solaiman, *The gradient of generative AI release: Methods and considerations.* Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, 2023. *https://arxiv.org/abs/2302.04844*

[24] John Kirchenbauer, *et al., A watermark for large language models,* 2023, arXiv preprint, *https://arxiv.org/pdf/2301.10226.pdf*

seen in recently launched Meta AI assistants)[25], without regular reinforcement or reminders that this is a persona. My answers to Questions 1 and 2 (Senator Schatz) provide examples of legislative initiatives regarding the disclosure of generative AI content and reasons for providing consumers with this information.

Trusted Provenance:

Question 3. In some contexts, like public safety, we may want a very secure label that clearly shows where a piece of content came from. What technical, policy, and educational steps are needed to increase the public's trust in the origin of critical information?

Answer. From a technical point of view, there are different approaches to signal the provenance of a piece of content. In my written statement I explain in further detail how invisible watermarking, cryptographic signatures and labeling can help increase the public's trust in information–and their limitations to each of them. A robust approach intended for specific public safety usages, for example critical government communications, would need to utilize a combination of these approaches to ensure redundancy and to enable a range of stakeholders (with varying capacities) to be able to confirm the integrity of the content. This approach also would likely require a pipeline of responsibility across foundation models, developers and deployers of AI models, as well as the creators and distributors of content to ensure these indications are durable and maintained.

In terms of policy steps, legislative proposals should avoid introducing a blanket requirement for compulsory disclosure of audiovisual AI. Hence, what constitutes 'public safety' should be clearly delineated and not used to undermine freedom of speech, deter technology whistleblowing, or enable surveillance. Any regulatory proposition should also unambiguously protect the right to privacy, and prohibit the default collection of personally identifiable information. For audiovisual content that is not produced via AI, legislators should steer clear from legally requiring labeling, but could consider how specific governmental contexts, for example, public safety agencies, utilize these disclosure mechanisms systematically and transparently.

Concerning education, media literacy will need to accompany the development of tools and legislation. It will be essential in order to enable consumers to understand the meaning of a label and verifiable provenance, and most importantly, comprehend how this information provides a signal of trust but does not equate to truth.[26] Relevant research and experience on consumer communication around tamper-proofing and accessible disclosure that exists in fields including currency, medicine and food safety could be relevant here. A specific usage in a particular public safety context of a set of provenance and disclosure indicators could also be a useful way to introduce the public to the concept and approach, rather than an attempt to do this at a more broad level.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. PETER WELCH TO SAM GREGORY

Question 1. AI systems are trained on massive datasets, from readily available information online to personal data collected across the internet. Generally, a larger training set corresponds with a more powerful AI tool, incentivizing AI developers to collect and exploit user data to strengthen their models. This raises significant concerns about protecting consumers' sensitive personal data and privacy.

a. Would you support legislative efforts to require platforms to allow users to opt-out of having their personal or service-generated data collected to train AI models?

b. How could these requirements benefit consumers, and what specific requirements should Congress consider as it explores providing consumers with the power to opt-out?

c. What additional steps should Congress take to empower consumers to protect their sensitive data from AI?

Answer. Users should have control over how their personal data, or data generated while interacting with a service, is used in the training of AI systems. This is consistent with a rights-based approach to regulating AI that emphasizes rights such as privacy. I have highlighted in my written statement that a rights-based approach to AI regulation, as opposed to exclusively a risk-based approach, is funda-

---

[25] Alex Heath, *Meta is putting AI chatbots everywhere,* The Verge, 27 September 2023, *https://www.theverge.com/2023/9/27/23891128/meta-ai-assistant-characters-whatsapp-instagram-connect*

[26] *Raquel Vazquez Llorente, Trusting Video in the Age of Generative AI, Common Place,* 01 June 2023, *https://commonplace.knowledgefutures.org/pub/9q6dd6lg/release/2*

mental to reflect the human rights impact of AI. Legislative efforts that protect users from having their data collected to train AI models would also be aligned with existing data protection laws in the U.S.[27]

Following a rights-based approach that centers U.S. Constitutional rights and fundamental human rights would mean to protect 'by default' these rights. Accordingly, users should have the choice to indicate whether they want their data included in the training—instead of having this information collected as a feature of the system, even with the possibility to opt-out. Government testing with 'nudging' techniques also suggests that in order to protect consumers' rights, it would be more desirable in this situation to give users the choice to opt-in.[28]

Additionally, the incorporation of 'do not train' tags could provide more nuanced control to creators and consumers over how their data is used. Approaches to media provenance and disclosure such as the C2PA have been considering how to incorporate these into media provenance, and provide creators and consumers with flexibility on how to include these markers. However, it is important to bear in mind that tag-based approaches are voluntary, and require companies and others engaged in AI-training to respect (or be compelled to oblige) the users' choice. There are other options that would be worth for Congress to explore. For instance, techniques based on adversarial attacks such as Glaze,[29] alter an image in a way that 'tricks' AI learning models while keeping changes minimally visible to the human eye–reducing the utility of these images for AI synthesis. Nonetheless, it is worth noting that these approaches will likely require constant updating in order to respond to attempts at circumventing their effectiveness.

Lastly, as highlighted in my written statement, a Federal data protection law would be a robust foundation to respond to future shifts in the AI landscape. More generally, to enhance the effectiveness of any data protection measures, Congress could also support and fund research that explores how the U.S. government can best communicate to consumers matters related to data protection in relation to AI, for example in users' interactions with chat-bots, as discussed in my answers to Questions 1 and 2 (Senator Schatz) and Question 2 (Senator Hickenlooper).

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. AMY KLOBUCHAR TO ROB STRAYER

Artificial Intelligence (AI) and Elections: Artificial intelligence (AI) has the potential to completely upend campaigns and elections unless there are rules of the road in place. I introduced bipartisan legislation with Senators Hawley, Coons, and Collins to prohibit the use of deceptive AI-generated audio, images, or video in political ads because some of this content needs to be banned altogether–and I also introduced the *REAL Political Ads Act* with Senators Booker and Bennet, which is led in the House by Rep. Yvette Clarke (D–NY), to require disclaimers on political ads containing AI-generated images or video so voters are aware when this technology is used in our elections.

Question 1. Do you agree that without giving people information to determine whether an image or video is created by AI, that generative AI poses a risk to our ability to hold free and fair elections?

Answer. Yes. Combatting deception stemming from the use of generative AI is a serious risk that must be mitigated by the private sector, media organizations, fact checkers, researchers, and consumers. Voters should know when they are interacting with a campaign ad that was generated by an AI system or product. To help address such risks, industry groups are working on efforts to develop digital content provenance and authentication techniques, including progress on a standard developed by the Coalition for Content Provenance and Authenticity (C2PA) to which ITI members are actively contributing.

Question 2. As AI makes spreading deceptive content cheaper and easier, what steps do you think should be taken to protect our democracy?

---

[27] For instance the Privacy Act of 1974; Health Insurance Portability and Accountability Act; Gramm-Leach-Blily Act; Children's Online Privacy Protection Act; and the California Consumer Privacy Act.

[28] On data protections and nudging, see: Acquisti, Alessandro, *et al., Nudges for privacy and security: Understanding and assisting users' choices online.* ACM Computing Surveys (CSUR) 50.3 (2017): 1–41; Jan M. Bauer, Regitze Bergstrøm, and Rune Foss-Madsen. *Are you sure you want a cookie?—The effects of choice architecture on users' decisions about sharing private online data.* Computers in Human behavior 120 (2021): 106729. Richard H. Thaler and Cass R. Sunstein, *Libertarian Paternalism,* Am Econ Rev. 93: (2003) 175–179.

[29] See: *https://glaze.cs.uchicago.edu/*

Answer. Developing industry standards for identifying AI generated images and content is a critical first step for providing meaningful disclosure. As discussed above, ITI member companies are heavily engaged in work to advance these efforts. ITI is also eager to work with policymakers to promote AI literacy, so that citizens can use the information provided by disclosure or transparency mechanisms.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. BRIAN SCHATZ TO ROB STRAYER

Question 1. Do people have the right to know when they are viewing AI-generated content or when they are interacting with an AI-chatbot?

Answer. Yes, ITI and our members believe consumers should be made aware through clear, concise, and conspicuous notifications when interacting with AI chatbots. Regarding AI-generated content, we believe disclosure requirements are warranted when such content has the potential to materially deceive the user.

Question 2. What are the dangers if individuals do not have a clear understanding that they are viewing or interacting with AI-generated content?

Answer. One of the primary concerns regarding AI-generated content that is not clearly labeled as such is that it is more difficult for individuals to discern what is real and what is not. This can serve to erode trust in the information individuals come across online.

Question 3. Do people have the right to know what model created the AI-generated content they are viewing or interacting with? Should generative AI systems affix information to AI-generated content reflecting that the content is generated by AI? Should that information be permanently affixed to the extent doing so is technically feasible?

Answer. The public deserves to know when they are interacting with AI-generated content that has the potential to materially deceive them. In certain high-risk situations, disclosure should be permanently affixed to the extent technically feasible. More research and technical work is needed to realize these goals and ensure labels, watermarks, or other types of content authentication tools, cannot be tampered with or otherwise removed.

Question 4. What role should platforms play in detecting AI-generated content on their sites and in informing consumers when content is AI-generated content?

Answer. Digital platforms have a significant role in detecting and informing consumers when AI-generated content is published. However, it is important to account for the challenges platform companies may experience in trying to authenticate AI-generated content uploaded by unknown and surreptitious users. AI authentication, which refers to determining when content is generated by AI technologies or verifying the provenance of digital layers therein, is a new and emerging field. Additional investment and research are required by stakeholders across the AI value chain, including platform companies, to determine which authentication techniques, such as provenance tracking, watermarking, metadata auditing, and/or human verification, are needed to limit the spread of mis- and dis-information and other concerning content. Importantly, ITI believes that a combination of the aforementioned methods are the most effective means to validate or authenticate AI-generated content.

RESPONSE TO WRITTEN QUESTION SUBMITTED BY HON. BEN RAY LUJÁN TO ROB STRAYER

Question. How can the U.S. government use responsible AI guardrails as a means to spur innovation and show the world that American AI systems are the most effective and trustworthy in the world?

Answer. The Administration, Congress, the private sector and other stakeholders will play an essential role in collaborating to advance guardrails that protect AI consumers and business users. It is the role of Congress to ensure legislation encourages future innovation and investment in the United States, protects consumers and businesses, and mitigates foreseeable risks. As such, the Federal Government should engage regularly with industry, academic, and civil society stakeholders to examine domestic and international standards and laws, identify policy gaps that warrant legislation, and support U.S. leadership in AI.

Looking at supply chain security, AI technologies are integral to the global information communications technology sector. Risk-based guardrails that focus on specific uses of AI in specific environments, when advanced in cooperation with U.S. allies and partners, can prevent the malicious use of AI and support future AI inno-

vation. Such an approach would contrast with unilateral, overly prescriptive regulations that may incentivize innovators to move outside the United States or regulations that designate all uses of AI in specific sectors and technologies as high-risk.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY JOHN HICKENLOOPER TO ROB STRAYER

International AI Regulation:

Question 1. The international community is closely watching what guardrails the United States establishes on high-risk AI systems. At the same time, the international community is moving forward with their own versions of legislation or frameworks to promote responsible use and development of AI systems. For example:

- The European Union is moving forward on passing their AI Act
- The United Kingdom is hosting an inaugural AI Summit this November focused on AI safety
- The United Nations and OECD are also moving forward with AI safety principles

Could you discuss the importance of the U.S. demonstrating international leadership on AI and building international consensus on our principles with our international allies?

Answer. It is imperative for the U.S. to advance international consensus on overarching principles to govern the development and use of AI and to remain a global leader in this space. The U.S. can remain a global leader by developing a vision for AI development and deployment in the United States, which includes taking a thoughtful, targeted approach to policymaking that is grounded in international, consensus-based standards, while also collaborating with partners around the world to ensure that any approaches to AI governance are interoperable and aligned to the extent possible. The potential for regulatory divergence is great, especially if countries undertake the development and deployment of AI in a vacuum. The Global Partnership on AI and the G7 Hiroshima AI Process have the potential to yield important principles and guidelines for advanced AI systems, and the U.S. should remain engaged in these and other multilateral efforts to help produce a vision of responsible AI.

Federal Capabilities:

Question 1. Several Federal agencies and offices have launched initiatives to advance safe and trustworthy AI, including the Department of Commerce, Department of Energy, and the Federal Trade Commission. In your view, which Federal agencies are equipped to implement aspects of AI policy? What strengths do these agencies offer to support trustworthy AI development or deployment, respectively?

Answer. Consistent with ITI's written testimony, before comprehensive AI legislation is introduced that governs the development and deployment of AI technologies to consumers and enterprises, Congress should examine the current legal and regulatory landscape to identify policy gaps that warrant legislation. Several Federal agencies already possess sufficient legal authority and jurisdiction to implement AI policy based on public input from the stakeholder community. For example, ITI has submitted feedback to multiple requests for public comments by the Department of Commerce, including its different subagencies and bureaus, and the Office of Science and Technology Policy (OSTP) in the White House, including:

- NIST AI Risk Management Framework: *Request for Information, Concept Paper, First Draft,* and *Second Draft*
- *NTIA Request for Comments on AI Accountability Policy*
- Misc.: *OSTP Request for Information on National Priorities for AI*

Consumer Notice:

Question 2. Generative AI systems are becoming more prevalent in the form of customer support chatbots and personal assistants. Should consumers be informed whether they are interacting with a generative AI system? What is the most effective way to plainly and easily notify consumers of this interaction?

Answer. Yes, ITI and our members believe consumers should be made aware through clear, concise, and conspicuous notifications when interacting with generative AI systems that have the potential to materially deceive or influence a user's decision-making. There is no one-size-fits-all approach for how an AI interface can plainly and easily notify consumers, but some examples include push notifications,

pop-up screens that grant consumers access to websites, and other clearly written disclaimers. The approach should consider whether the generative AI system materially impacts an individual's access to goods and services.

Review Boards:

Question 3. Some companies have established Internal Boards to assess and determine whether an AI system is developed with safety, risk mitigation, and ethics in mind. The majority of Internal Boards consist of personnel within the company, and may not always include independent experts. In general, what is the typical review process for an AI system conducted by engineers or an Internal Board within a member company? If engineers or Internal Board members draw different conclusions after an AI system review, how are disputes resolved among these parties?

Answer. Internal ethics boards are another mechanism employed by ITI companies to instill and foster greater trust in AI solutions before such technologies are placed into the market. The review process and board composition, understandably, will vary by company, but the general intent is to organize a review process that collects diverse perspectives from experts that play an essential role in overseeing the development of AI technologies.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. PETER WELCH TO ROB STRAYER

Question 1. Information privacy protections have not necessarily kept pace with emerging AI technologies, posing particular risk in the health care space. Evidence shows that biases often exist in AI diagnostic technology—for example, in the rate of detection of breast cancers for Black Americans versus White Americans. These tools are often not reliable and more importantly, they aren't subject to a universal evaluation standard.

a. What are the biggest risks that the expanded use of AI poses in the health care sector, particularly as it relates to patient outcomes, diagnostic tools, and bias? What steps should the industry and Congress take to mitigate these risks?

b. What specific privacy risks do AI tools pose for patients in the health care space?

c. What steps should Congress consider to bolster patient privacy protections as AI tools become more common in the health sector?

d. As Congress works to minimize potential harms, what steps should we take to ensure we do not stifle AI innovations that would benefit patients?

Answer. Regarding privacy protections, ITI broadly believes that U.S. consumers, including medical patients, should be in full control of how their personal data is collected, transmitted and stored. Individuals should have the right to exercise control over the use of their personal data where reasonable to the context surrounding the use of personal data. These individual control rights, consistent with the rights and legal obligations of other stakeholders, include the legal requirements set forth in HIPAA that pertain to the right to access, correct, port, delete, consent to, and object to the use of personal data about themselves. For additional information on ITI's viewpoints on consumer data privacy, we encourage staff to review ITI's *Framework to Advance Interoperable Rules (FAIR) on Privacy.*

We take seriously concerns related to biased outcomes and/or biased AI systems. Our members are committed to taking steps to mitigate bias, including conducting impact and risk assessments throughout the development and deployment lifecycle and coordinating with NIST to identify standards and steps needed to mitigate bias. ITI supports NIST's efforts to develop a standard aimed at identifying and mitigating bias (*SP–1270*), as this work will be important to bolster industry efforts. We encourage Congress to continue to support this work, as measurement techniques will be especially important to consistently identifying and managing bias. Bias can be introduced at different points throughout the AI lifecycle, so an effective approach to mitigating bias requires a holistic approach that considers risk of bias throughout.

Question 2. Building a generative AI model requires intense computational and financial resources. This barrier has resulted in Big Tech firms—such as Google, Microsoft, and Meta—leading in the development and release of AI products. Big Tech's dominance in data gives these companies access to behavioral data insights that provide them with competitive advantages compared to smaller, emerging tech companies. These same Big Tech firms have made dozens of strategic acquisitions and investments in foundational AI tech to gain access to new technologies and capabilities, or potentially foreclose on rivals.

a. Do you believe AI will increase competition, or will it lead to further consolidation within the Big Tech landscape?

b. What guardrails should Congress put in place to ensure the development of generative AI will enhance competition within the tech marketplace?

Answer. ITI members welcome increased competition across the AI ecosystem, which extends to AI startups and small, medium, and large enterprises. While ITI primarily represents large technology sector companies, our members rely heavily on partnerships with AI startups and utilizing small businesses as a customer base for future growth and economic opportunities. Additionally, Congress should consider legislation that encourages strategic partnerships between large enterprises, AI startups, and other entities and institutions that prioritize R&D projects.

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. TED CRUZ TO ROB STRAYER

Question 1. ***Regulation.*** As I noted in my opening remarks for the record, the Biden Administration and some of my colleagues in Congress have—unfortunately—embraced doomsday AI scenarios to justify expanded Federal powers. Some of these proposals are extremely onerous: licensing regimes, creating a new regulatory agency to police computer code, and mandatory, NEPA-style impact assessments before an AI model can be used.

These types of regulations could cripple innovation and put us at a competitive disadvantage against AI advancements in places like China. Just like with past stages of tech, the best ideas often come from individuals and small startups. There is no question that new AI regulation will target the ideas and intellectual property of individuals. Smaller companies with fewer resources would face disproportionately large compliance burdens that would be expensive, confusing, and time-consuming.

Large companies with more resources may be better able to weather such regulations. I'm concerned that large companies are dominating the conversation about AI regulations. For example, no small companies participated in Senator Schumer's first AI "Insight Forum" on September 13. Similarly, we have only seen large AI players sign onto the White House's "voluntary" commitments on AI regulation.

a. How are smaller businesses reacting to the recent push to impose regulations on AI in the U.S., as well as in the EU?

Answer. ITI members welcome increased competition across the AI ecosystem which extends to AI startups and small, medium, and large enterprises. While ITI primarily represents large technology sector companies, our members rely heavily on partnerships with AI startups and utilizing small businesses as a customer base for future growth and economic opportunities. Additionally, Congress should consider legislation that encourages strategic partnerships between large enterprises, AI startups, and other entities and institutions that prioritize R&D projects. Haphazard or overly broad regulations will negatively impact companies of all sizes, especially small businesses, so it is imperative that Congress be equipped with the tools and expertise needed to advance meaningful legislation that encourages future AI innovation and investment in the United States while mitigating real risks to consumers and businesses alike.

b. Can you please share some of their primary concerns with these regulations and describe how regulations would affect innovation in the AI field?

Answer. In general, regulatory approaches that leverage an overly broad definition of AI, that do not take a risk-based approach, and that mandate overly prescriptive requirements will stymie innovation. Regulation that designates entire classes of technology as high-risk, or that designates entire sectors of the economy as high-risk, misses important nuance and may implicate many low-risk uses of the technology. Additionally, regulation that provides market surveillance authorities access to source code could serve to significantly impede innovation.

For example, although negotiations continue in the European Union on its AI Act, we remain concerned with the direction of certain provisions, especially those concerning foundation models and/or general purpose AI (GPAI) systems. Presently, both the European Parliament and the European Council text seek to regulate the development of foundation models and/or GPAI systems, applying high-risk obligations to this class of AI technology. However, it is our view that a risk-based approach should also apply to this type of AI technology, and that further nuance is necessary in the definitions of both foundation models and GPAI system to clearly differentiate between the two. Only where a foundation model is implemented in a high-risk use case should it be subject to the requirements of the AI Act. Impor-

tantly, we are supportive of provisions that ensure a balanced allocation of responsibilities across the value chain so that those who implement a foundation model into a high-risk use case can comply.

For additional information, please review our recent blog posts on the subject: *The EU AI Act: What it Means for Global Policymaking* and *Five Recommendations for the EU AI Act Trilogue Negotiations.*

Question 2. ***China.*** I am deeply concerned that if we hamper innovation in the United States, we will allow China to get ahead of us. Xi Jinping views AI as key for China's military capabilities, and the Chinese Communist Party (CCP) will stop at nothing to accelerate their development and adoption. They are not holding themselves back on questions of ethics. While we want to be responsible, we need to be realistic. We need to accept some level of risk to ensure American innovation thrives.

a. What happens if the United States fails to innovate more quickly than China?

Answer. China has a well-established record of shifting the playing field in its favor—whether it is creating conditions for technology transfer through forced partnerships with Chinese companies; establishing ambiguous and intrusive security review regimes; or circumventing U.S. export controls laws, these unfair practices not only create an unfair economic advantage, but may also, especially in the case of AI innovation, pose a national security risk. Regardless of whether China plays by the rules or not, China will continue to invest and improve in AI technological development, innovation, and growth.

If the United States were to cede leadership in AI innovation to China, and by extension the CCP, China will undoubtedly leverage its state-owned enterprises and technology-driven initiatives, such as the *Belt and Road Initiative,* to deploy AI technology across China's global sphere of influence, which could establish a standard of practice abroad that weakens fundamental human rights and exacerbates pervasive mass-surveillance practices. Another example, according to the *Department of Homeland Security's 2024 Threat Assessment,* warns that China will likely deploy novel technologies such as AI-developed malware to bolster the effectiveness of cyber operations and enable larger scale attacks that threaten U.S. critical infrastructure.

○