

TRUSTWORTHY AI: MANAGING THE RISKS OF ARTIFICIAL INTELLIGENCE

HEARING BEFORE THE SUBCOMMITTEE ON RESEARCH AND TECHNOLOGY OF THE COMMITTEE ON SCIENCE, SPACE, AND TECHNOLOGY OF THE HOUSE OF REPRESENTATIVES ONE HUNDRED SEVENTEENTH CONGRESS SECOND SESSION

SEPTEMBER 29, 2022

Serial No. 117-70

Printed for the use of the Committee on Science, Space, and Technology



Available via the World Wide Web: <http://science.house.gov>

U.S. GOVERNMENT PUBLISHING OFFICE

48-617PDF

WASHINGTON : 2023

COMMITTEE ON SCIENCE, SPACE, AND TECHNOLOGY

HON. EDDIE BERNICE JOHNSON, Texas, *Chairwoman*

ZOE LOFGREN, California	FRANK LUCAS, Oklahoma,
SUZANNE BONAMICI, Oregon	<i>Ranking Member</i>
AMI BERA, California	MO BROOKS, Alabama
HALEY STEVENS, Michigan,	BILL POSEY, Florida
<i>Vice Chair</i>	RANDY WEBER, Texas
MIKIE SHERRILL, New Jersey	BRIAN BABIN, Texas
JAMAAL BOWMAN, New York	ANTHONY GONZALEZ, Ohio
MELANIE A. STANSBURY, New Mexico	MICHAEL WALTZ, Florida
BRAD SHERMAN, California	JAMES R. BAIRD, Indiana
ED PERLMUTTER, Colorado	DANIEL WEBSTER, Florida
JERRY MCNERNEY, California	MIKE GARCIA, California
PAUL TONKO, New York	STEPHANIE I. BICE, Oklahoma
BILL FOSTER, Illinois	YOUNG KIM, California
DONALD NORCROSS, New Jersey	RANDY FEENSTRA, Iowa
DON BEYER, Virginia	JAKE LATURNER, Kansas
SEAN CASTEN, Illinois	CARLOS A. GIMENEZ, Florida
CONOR LAMB, Pennsylvania	JAY OBERNOLTE, California
DEBORAH ROSS, North Carolina	PETER MEIJER, Michigan
GWEN MOORE, Wisconsin	JAKE ELLZEY, TEXAS
DAN KILDEE, Michigan	MIKE CAREY, OHIO
SUSAN WILD, Pennsylvania	
LIZZIE FLETCHER, Texas	
VACANCY	

SUBCOMMITTEE ON RESEARCH AND TECHNOLOGY

HON. HALEY STEVENS, Michigan, *Chairwoman*

MELANIE A. STANSBURY, New Mexico	RANDY FEENSTRA, Iowa,
PAUL TONKO, New York	<i>Ranking Member</i>
GWEN MOORE, Wisconsin	ANTHONY GONZALEZ, Ohio
SUSAN WILD, Pennsylvania	JAMES R. BAIRD, Indiana
BILL FOSTER, Illinois	JAKE LATURNER, Kansas
CONOR LAMB, Pennsylvania	PETER MEIJER, Michigan
DEBORAH ROSS, North Carolina	JAKE ELLZEY, TEXAS

C O N T E N T S

September 29, 2022

	Page
Hearing Charter	2
Opening Statements	
Statement by Representative Haley Stevens, Chairwoman, Subcommittee on Research and Technology, Committee on Science, Space, and Technology, U.S. House of Representatives	9
Written Statement	10
Statement by Representative Randy Feenstra, Ranking Member, Subcommittee on Research and Technology, Committee on Science, Space, and Technology, U.S. House of Representatives	10
Written Statement	12
Written statement by Representative Eddie Bernice Johnson, Chairwoman, Committee on Science, Space, and Technology, U.S. House of Representatives	13
Witnesses:	
Ms. Elham Tabassi, Chief of Staff, Information Technology Laboratory, National Institute of Standards and Technology	
Oral Statement	14
Written Statement	17
Dr. Charles Isbell, Dean and John P. Imlay, Jr. Chair of the College of Computing, Georgia Institute of Technology	
Oral Statement	28
Written Statement	30
Mr. Jordan Crenshaw, Vice President of the Chamber Technology Engagement Center, U.S. Chamber of Commerce	
Oral Statement	36
Written Statement	38
Ms. Navrina Singh, Founder and Chief Executive Officer, Credo AI	
Oral Statement	49
Written Statement	51
Discussion	61
Appendix I: Answers to Post-Hearing Questions	
Ms. Elham Tabassi, Chief of Staff, Information Technology Laboratory, National Institute of Standards and Technology	86
Mr. Jordan Crenshaw, Vice President of the Chamber Technology Engagement Center, U.S. Chamber of Commerce	87
Appendix II: Additional Material for the Record	
Document submitted by Representative Brad Sherman, Committee on Science, Space, and Technology, U.S. House of Representatives	
“Engineered Intelligence: Creating a Successor Species,” Representative Brad Sherman	92

TRUSTWORTHY AI: MANAGING THE RISKS OF ARTIFICIAL INTELLIGENCE

THURSDAY, SEPTEMBER 29, 2022

HOUSE OF REPRESENTATIVES,
SUBCOMMITTEE ON RESEARCH AND TECHNOLOGY,
COMMITTEE ON SCIENCE, SPACE, AND TECHNOLOGY,
Washington, D.C.

The Subcommittee met, pursuant to notice, at 10:42 a.m., in room 2318, Rayburn House Office Building, Hon. Haley Stevens [Chairwoman of the Subcommittee] presiding.

**U.S. HOUSE OF REPRESENTATIVES
COMMITTEE ON SCIENCE, SPACE, AND TECHNOLOGY
SUBCOMMITTEE ON RESEARCH AND TECHNOLOGY
HEARING CHARTER**

Trustworthy AI: Managing the Risks of Artificial Intelligence

**Thursday, September 29, 2022
10:30 am – 12:30 pm
2318 Rayburn House Office Building and Online via Zoom**

PURPOSE

On Thursday, September 29, 2022, the Subcommittee on Research and Technology of the Committee on Science, Space, and Technology will hold a hearing to discuss tools, best practices, and challenges in the design, development, testing, and deployment of trustworthy artificial intelligence (AI) systems. The Subcommittee will examine efforts in academia, industry, and government to create a culture of responsibility around AI systems, identify and remove harmful bias in AI systems, improve explainability and transparency of AI systems, and mitigate other risks associated with AI systems. The Subcommittee will also explore the National Institute of Standards and Technology's ongoing efforts to create an artificial intelligence risk management framework.

WITNESSES

- **Ms. Elham Tabassi**, Chief of Staff, Information Technology Laboratory, National Institute of Standards and Technology
- **Dr. Charles Isbell**, Dean and John P. Imlay, Jr. Chair of the College of Computing, Georgia Institute of Technology
- **Mr. Jordan Crenshaw**, Vice President of the Chamber Technology Engagement Center, U.S. Chamber of Commerce
- **Ms. Navrina Singh**, Founder and Chief Executive Officer, Credo AI

OVERARCHING QUESTIONS

- What are the risks that can arise from the development and deployment of AI systems, including how harmful biases can arise in these systems?
- What are the activities being undertaken by academia, industry, and the government to develop, test, and responsibly deploy trustworthy AI systems?
- How should the United States encourage more organizations to think critically about risks that arise from AI systems, including at the earliest stages of development?
- Where should the Federal government focus efforts to promote the development and deployment of trustworthy artificial intelligence across every sector of the economy?

BACKGROUND

Artificial intelligence refers to the theory and development of computer systems that can perform tasks that would normally require human intelligence, such as decision making or speech recognition. Modern AI systems are engineered or machine-based systems that can, for a given set of human-defined objectives and with varying levels of autonomy, generate predictions, recommendations, or decisions

influencing real or virtual environments.¹ All applications of artificial intelligence in use today can be considered “narrow AI,” or AI that is designed to do a very specific set of tasks. In contrast, artificial general intelligence is a theoretical system that possesses generalized human cognitive abilities and, when presented with an unfamiliar and complex problem, could develop solutions drawing from contextual knowledge. Modern systems are likely decades away from achieving artificial general intelligence.

Most AI systems are developed using a technique called machine learning, which involves developing an algorithmic model based on input data, then using that model to make certain optimizations or predictions. An example of this is image recognition, in which a set of human-labeled images (e.g., “traffic lights” in CAPTCHA tests that users take when logging into a website) are fed into an algorithm, which then looks for patterns common to all images with a specific label. The algorithm builds a model (i.e., “learns”) from this “training data”, so when it is presented with an unlabeled image containing one of the objects that was in the training data, it can make a guess as to what the object is. This method of training algorithms with human-labeled data is called “supervised learning”. There is also “unsupervised learning”, in which no labels are provided, and the algorithm simply looks for similarities and groups images into clusters based on certain characteristics. Additionally, there is “reinforcement learning”, in which an algorithm interacts with its environment, executes actions, and learns through trial and error.

While AI systems have been in use in the commercial sector for decades, recent advances in computing, improved software engineering, and better access to large data sets have markedly increased the capabilities of AI systems. As a result, AI systems have led to a wide range of innovations with the potential to benefit nearly all aspects of our society and support our economic and national security. AI systems are increasingly used in scientific research to help sort and analyze massive amounts of data in fields such as weather prediction, cosmology, and genetics research. Recent advances in natural language processing and image generation have led to AI systems that can write text or generate art.²

AI RISKS

While AI-systems have the potential to improve our lives, in sometimes transformative ways, they also have the potential to do significant harm if risks associated with these systems are not mitigated. While risks to any type of information-based system also apply to AI systems (e.g., privacy, cybersecurity, and safety concerns), these systems also create a set of risks that require specific consideration. AI systems can amplify, perpetuate, and exacerbate existing structural inequalities in our society, or create new ones. AI systems can also exhibit unintended properties with potential ethical, safety, or security consequences for individuals or communities. Risks associated with AI systems arise from the data used to train the AI system, the system itself, the use of the system, or interaction of people with the system. Importantly, AI systems and their associated risks are socio-technical, meaning they are a product of the complex human, organizational, and technical factors involved in their design, development, and use. For example, questions of fairness or equity caused by the decisions of AI systems relate to societal dynamics and human behavior. Purely technical solutions will not solve societal challenges.

Harmful Bias

One major set of risks caused by AI systems is harmful bias, which can occur when an algorithm produces results that are systemically prejudiced due to erroneous assumptions in the machine learning

¹ “AI Risk Management Framework: Second Draft,” [NIST](#), August 18, 2022.

² “GPT-3 Powers the Next Generation of Apps,” [OpenAI](#), March 25, 2021; “DALL·E: Creating Images from Text,” [OpenAI](#), January 5, 2021.

process. Bias can be introduced purposefully or inadvertently into an AI system, or it can emerge as the system is being deployed. For example, a facial recognition system trained mostly on light-skinned faces will perform poorly identifying faces with darker skin, and a facial recognition system trained to perfection in a lab may fail when encountering real-world scenarios. Moreover, intentional or unintentional changes during training may fundamentally alter AI system performance.

According to the National Institute of Standards and Technology (NIST), there are three categories of bias.³ First, systemic biases result when AI systems create advantages for certain social groups while disadvantaging others. Systemic bias is also referred to as institutional or historical bias. Systemic biases can creep their way into datasets or can be reinforced by institutional norms, practices, and processes across the AI lifecycle. Second, statistical and computational biases result from errors that occur due to a sample that the AI system is trained on not being representative of the population. These biases often arise when algorithms are trained on one type of data and cannot extrapolate beyond those data. Finally, human biases reflect systematic errors in human thought. These biases are often implicit and tend to relate to how an individual or group perceives information to make a decision or fill in missing or unknown information. Because AI systems are designed by humans, this type of bias is present across the entire AI lifecycle.

Not all bias is harmful. Statistical and computational biases that arise in an analysis are a normal part of data science. Bias can also be beneficial, such as algorithms that use data on an individual's habits to tailor new content based on their interests. However, many cases of bias can cause significant harm. For example, a self-driving car trained by driving on the roads of Boston may not recognize different patterns in other cities, and an AI diagnostic tool trained on x-ray images of younger patients may fail to perform well on older patients. Combatting harmful bias in AI will require better alignment between AI tasks and actual human goals. While it will require additional technology expertise to improve the detection and mitigation of bias, it will also require an understanding of the relevant social and ethical considerations.

Explainability and Interpretability

Some AI systems are functionally black boxes, which means it is difficult to understand why algorithms make the decisions that they do. For example, one type of machine learning system is called a "neural network," which consists of thousands or even millions of simple processing nodes that are densely interconnected. Training data is fed to the bottom layer and as it passes through the succeeding layers it gets multiplied and added together in complex ways, until it finally arrives at the output layer in its transformed final state. Due to the complexity, scientists are unable to fully understand these interactions in a useful way. Observers can only effectively assess this process by reviewing an algorithm's inputs and outputs.

This challenge has given rise to fields of research focused on assessing and understanding algorithmic decisions. For example, researchers and companies are working to improve algorithmic explainability, or the ability of algorithms to explain their decisions. However, modern explainability techniques come with trade-offs—improving the explainability of algorithms has often come at the cost of accuracy of outputs.⁴ In contrast, some researchers are focused on interpretability, which refers to techniques used to understand the meaning of AI systems' output in the context of its designed functional purpose. One area of focus for interpretability is called test, evaluation, validation, and verification (TEVV), which uses

³ Reva Schwartz et al., "Towards a Standard for Identifying and Managing Bias in Artificial Intelligence," [NIST](#), March 2022.

⁴ Cynthia Ruden, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," [Nature Machine Intelligence](#), vol. 1, 2019, 206–215.

separate AI actors to examine an AI system or its components or detect and remediate problems throughout the AI lifecycle.⁵

Safety

Ensuring AI systems are safe means preventing them from leading to physical or psychological harm, or creating a state in which human life, health, property, or the environment is endangered.⁶ One major challenge of AI safety is ensuring the system can continue to operate safely in unfamiliar situations. For example, modern autonomous vehicles can only operate in certain environments under certain conditions in a safe manner.⁷ Another challenge to achieving AI safety is avoiding misspecification, or poor alignment between an AI behavior and the system designer's intentions. Misspecification occurred in YouTube's video recommendation algorithm when an AI system that was optimized for user engagement unintentionally directed users to extremist content.⁸

Safety issues are mostly dealt with through careful design, planning, and testing to prevent failures, conditions, or environments in which it becomes dangerous to use an AI system. According to NIST, practical approaches to AI safety include "rigorous simulation and in-domain testing, real-time monitoring, and the ability to shut down or modify systems that deviate from intended or expected functionality".⁹

Cybersecurity and Privacy

While AI systems are susceptible to the same privacy and security risks as all information-based systems, there are some concerns that are unique to AI systems. AI systems have more complex attack surfaces that can enable malicious actors to compromise their security more easily. For example, malicious actors could theoretically make alterations to open-source datasets to manipulate an AI system to produce an inaccurate or harmful result.¹⁰ Similarly, AI systems could be trained outside an organization's security controls or trained in one domain and then "fine-tuned" for another, resulting in vulnerabilities. As a result, existing privacy and cybersecurity guidance are ill-equipped to ensure the data protection of AI systems.

Computational Costs

Training AI systems requires a large amount of computational power. Since 2012, the amount of computational power used to train the largest AI systems has been increasing exponentially—doubling every 3.4 months.¹¹ A paper in 2019 found that training a single large-scale AI system required five times as much carbon as the lifetime emissions of the average American car.¹² If the United States is to avert the climate crisis while maintaining its global leadership in AI, the research community and tech industry should explore more efficient AI training methodologies and more efficient computing systems.

⁵ "AI Risk Management Framework: Second Draft," [NIST](#).

⁶ Ibid.

⁷ Several automakers have achieved level four automation in their vehicles. See "Levels of Automation" [NHTSA](#), accessed September 22, 2022.

⁸ Homa Hosseinmardi et al., "Examining the consumption of radical content on YouTube," Complex Networks & Their Applications, hosted on [Proceedings of the National Academy of Sciences](#), 2022, 166-177.

⁹ "AI Risk Management Framework: Second Draft," [NIST](#).

¹⁰ Andrew Lohn, "Poison in the Well," [Center for Security and Emerging Technology](#), June 2021.

¹¹ Jack Clark, "AI and Compute," [OpenAI](#), May 16, 2018.

¹² Emma Strubell et al., "Energy and Policy Considerations for Deep Learning in NLP," In the 57th Annual Meeting of the Association for Computational Linguistics (ACL), [stored in arxiv](#), July 2019.

GOVERNMENT ACTION

In December 2020, Congress enacted the *National Artificial Intelligence Initiative Act* or NAIIA (P.L. 116-283). This bipartisan legislation, which was led by the House Science Committee, accelerated and coordinated Federal investments and new public-private partnerships in research, standards, and education in trustworthy artificial intelligence. The law establishes interagency coordination and strategic planning efforts in AI research, development, standards, and education through an Interagency Coordination Committee and a coordination office managed by the Office of Science and Technology Policy (OSTP). The legislation also created the National AI Advisory Committee (NAIAC) to assess the implementation of the law, track advancements in AI science, and propose recommendations to advance U.S. competitiveness in AI. The Department of Commerce selected members for the NAIAC in May 2022, with the plan to publish a report in 2023.¹³ Finally, the legislation directed the Department of Energy (DOE), the National Science Foundation (NSF), and Department of Commerce research agencies to conduct AI-related activities, many of which are designed to assess and mitigate AI-related risks.

OSTP

OSTP has pursued several initiatives related to promoting trustworthy AI. In 2021, OSTP announced an effort to develop a bill of rights for an automated society, also called the “AI bill of rights”.¹⁴ OSTP has sought input from the boarder community on what this document should contain. In March 2022, OSTP also sought feedback on updating the National AI Research and Development Strategic Plan, which includes strategic aims to both “understand the ethical, legal, and societal implications of AI” and “ensure the safety and security of AI systems”.¹⁵

National Institute of Standards and Technology

NIST, which is housed within the Department of Commerce, conducts fundamental and applied research and measurement activities to cultivate trust and improve the design, development, and governance of AI systems. NIST published principles of explainable AI in 2020 before NAIIA was enacted.¹⁶ In NAIIA, Congress directed NIST to expand upon these efforts by developing a voluntary AI risk management framework through collaboration with stakeholders across public and private sectors. To date, NIST has held two workshops to develop the AI risk management framework, released two drafts of the framework, and published a draft playbook to help with implementation.¹⁷ NIST plans to publish the first version of the AI risk management framework in January 2023.

In addition, NIST is conducting several other trustworthy AI-related activities, including:

- Developing taxonomy, terminology, and testbeds for measuring risks in AI systems and informing the standards needed for key technical characteristics of AI trustworthiness.
- Developing data characterizations, key practices for data documentation, and datasets that the broader community can use to test or train AI systems while preserving privacy and cybersecurity.

¹³ “Commerce Department Launches the National Artificial Intelligence Advisory Committee,” [Department of Commerce](#), May 4, 2022.

¹⁴ “Join the Effort to Create A Bill of Rights for an Automated Society,” [White House](#), November 10, 2021.

¹⁵ Office of Science and Technology Policy, “Request for Information to the Update of the National Artificial Intelligence Research and Development Strategic Plan,” [Federal Register](#), February 2, 2022.

¹⁶ P. Jonathon Phillips et. al., “Four Principles of Explainable Artificial Intelligence,” [NIST](#), September 2021.

¹⁷ “AI Risk Management Framework: Second Draft,” [NIST](#).

- Coordinating across the government and with industry stakeholders to identify critical standards development activities, strategies, and gaps for trustworthy AI.¹⁸
- Developing guidance to facilitate voluntary data sharing arrangements among industry, federally funded research centers, and federal agencies to advance AI research and technologies.

National Science Foundation

Achieving the responsible design and deployment of AI also requires integrating ethics into technology education and research at every stage—from K-12 education to AI developers. It requires viewing AI as an interdisciplinary field rather than a purely technical field. NSF funds university research across all non-biomedical disciplines (including social sciences) and numerous STEM education programs. As a result, the agency will play a key role in achieving these goals. In NAIIA, Congress directed NSF to make awards supporting research that contributes to the development of trustworthy AI, supports K-12, undergraduate, and graduate education on trustworthy AI, and creates faculty technology ethics fellowships to support more research into the field of technology ethics.¹⁹ Moreover, the *CHIPS and Science Act of 2022* (P.L. 117-167) directs NSF to establish a requirement for an ethics statement in award proposals to ensure researchers are considering the social implications of their work.²⁰ NSF also funds a network of 18 AI research institutes, each devoted to a different sector or AI-related challenge. This combined investment of \$220 million reaches a total of 40 states and the District of Columbia. In 2021, NSF announced a partnership with NIST to establish an AI research institute on trustworthy AI.²¹ The winner of this solicitation should be announced later this year.

International

There are also international conversations taking place surrounding the development and responsible deployment of trustworthy AI. The Organisation for Economic Cooperation and Development (OECD) adopted a set of AI principles for guiding governments in responsible stewardship of trustworthy AI in 2019.²² Many individual countries have also established their own AI strategies that incorporate ethics to various extents. Singapore was one of the first to develop an AI governance framework in 2019, later iterations of which evolved into a practical toolkit for companies to demonstrate trustworthy AI in a practical manner.²³ The European Union proposed the AI Act in 2021 to harmonize regulations as they relate to AI systems, including a process for self-certification and government oversight of many categories of high-risk AI systems.²⁴ Because AI risk management is a relatively new activity and organizations are required to self-certify that they control for AI-risks, there is significant uncertainty surrounding the pending EU law's requirements. Many U.S. companies are looking to the NIST AI risk management framework as a possible solution to this dilemma.

PRIVATE SECTOR ACTION

The private sector is also attempting to tackle issues related to developing and deploying trustworthy AI. Companies such as Microsoft, Google, and Intel have all published their own versions of AI ethics

¹⁸ “U.S. LEADERSHIP IN AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools,” [NIST](#), August 9, 2019.

¹⁹ Rep. Eddie Bernice Johnson, *The National AI Initiative Act*, H.R. 6216, incorporated into [H.R. 6395](#), 116th Cong.

²⁰ Rep. Eddie Bernice Johnson, *CHIPS and Science Act of 2021*, H.R. 4521, incorporated into [H.R. 4346](#), 117th Cong.

²¹ James Donlon and Rebecca Hwa, “National Artificial Intelligence (AI) Research Institutes,” [NSF](#), November 16, 2021.

²² “OECD AI Principles,” [OECD](#), February 2019.

²³ “Singapore’s Approach to AI Governance,” [Singapore Personal Data Protection Commission](#), May 25, 2022.

²⁴ “The AI Act” [European Union](#), April 2021.

principles.²⁵ Many industry groups are also engaging in their own activities to promote trustworthy AI development and deployment. For example, the U.S. Chamber of Commerce has launched a bipartisan commission on AI to “advance U.S. leadership in the use and regulation of AI technology.”²⁶ Many of these principles developed by industry are generally abstract and lack concrete governance structures and accountability measures. However, some major technology companies have begun to develop and implement concrete measures.²⁷ Other businesses are developing tools and practical methodologies to help organizations assess and mitigate AI-related risks. The Mozilla foundation is funding open-source AI auditing tools.²⁸ Some companies have developed proprietary tools that enable their clientele to identify and mitigate AI risks.²⁹

²⁵ “Responsible AI,” [Microsoft](#), accessed September 21, 2022; Responsible AI Practices,” [Google](#), accessed September 21, 2022; “Intel’s Recommendations for the U.S. National Strategy on Artificial Intelligence,” [Intel](#), March 5, 2019.

²⁶ “U.S. Chamber Launches Bipartisan Commission on Artificial Intelligence to Advance U.S. Leadership,” [U.S. Chamber of Commerce](#), January 18, 2022.

²⁷ Jon Belkowitz and Leah Koshiyama, “Trust in the Time of AI: Why Salesforce Invests in Ethical Guardrails,” [Salesforce](#), April 19, 2022.

²⁸ “Mozilla Technology Fund Seeks People, Projects Auditing AI Systems with Open-Source Approaches,” [Mozilla Foundation](#), September 6, 2022.

²⁹ For examples, please see [ORCAA](#) and [Credo AI](#).

Chairwoman STEVENS. Welcome to the Research and Technology hearing to examine the harmful impacts associated with artificial intelligence (AI) systems, as well as the opportunities with our artificial intelligence systems, the activities that academia, government, and industry are conducting to prevent, mitigate, and manage AI risks as these new technologies proliferate.

I'm thrilled to be joined by this distinguished panel of witnesses, all of whom are in the room with us today. It is great to see your faces and to be together the first time since a March 2020 hearing, I believe.

It is also of deep importance to be discussing the benefits and the challenges of artificial intelligence, the potential to influence many aspects of our lives and support our economic and national security. The applications in our everyday lives span from merely convenient like recommending your next movie, to transformational, like aiding doctors in earlier detection of disease. In my home State of Michigan, advances in artificial intelligence by automakers are accelerating the development of autonomous vehicles that will lead to reduced traffic and increased road safety. Artificial intelligence systems are also increasingly used to analyze massive amounts of data to propel research in fields to enhance our understanding of the universe and cosmology, to synthetic biology, to weather prediction. Call our ancestors.

But ill-conceived or untested applications of artificial intelligence have also on occasion caused damage. We have already seen ways AI systems can amplify, perpetuate, or exacerbate inequitable outcomes. Researchers have shown that AI systems making decisions in high-risk situations, such as credit or housing, can be biased against already disadvantaged communities, causing harm. This is why we need to encourage people developing or deploying AI systems to be thoughtful about what they're putting out into the world. We must develop the tools, methodologies, and standards to ensure that AI products and services are safe and secure, accurate, free of harmful bias, and otherwise trustworthy. We are in a moment of trust.

Since taking over this gavel of the Research and Technology Subcommittee a few years ago, I have worked with my colleagues on both sides of the aisle to promote trustworthy AI. We're working together. I was proud to secure trustworthy AI provisions in the *CHIPS and Science Act* that was passed and signed into law just last month, which also promotes the—or includes the *Promoting Digital Privacy Technologies Act*, which passed the House and awaits a vote in the Senate, supports privacy-enhanced data sets and tools for training AI systems.

Additionally, this Committee led the development of the 2020 *National AI Initiative Act* to accelerate and coordinate Federal investments in research standards and education of trustworthy AI. In that act we also directed NIST (National Institute of Standards and Technology) to develop an AI Risk Management Framework (AI RMF) to help organizations understand and mitigate the risks associated with these technologies.

We're all excited to be having today's hearing and to discuss the progress of this work and the many other things that NIST is doing to promote trustworthy AI. Academia and industry are supporting

ethical approaches to artificial intelligence. Universities across the country are adopting principles for responsible use of AI and incorporating ethics into their computer science (CS) curricula. Industry is moving past theoretical principles into practical approaches to mitigating AI risks. There's more to do, there's jobs to be had, and people's lives are being impacted.

With that, we're here in Congress to ensure that the United States continues to lead the world in artificial intelligence and trustworthy artificial intelligence. And we thank our witnesses for their time.

[The prepared statement of Chairwoman Stevens follows:]

Good morning and welcome to today's Research and Technology hearing to examine the harmful impacts associated with artificial intelligence systems, and the activities that academia, government, and industry are conducting to prevent, mitigate, and manage AI risks. I am thrilled to be joined by our distinguished panel of witnesses. It is great to be with you all in person today, and I look forward to hearing your testimony.

Artificial intelligence has the potential to benefit many aspects of our lives and support our economic and national security. The applications in our everyday lives span from merely convenient, like recommending your next movie, to transformational, like aiding doctors in earlier detection of disease. In my home state of Michigan, advances in AI by automakers are accelerating the development of autonomous vehicles that will lead to reduced traffic and increased road safety. AI systems are also increasingly used to analyze massive amounts of data to propel research in fields to enhance our understanding of the universe in cosmology to synthetic biology to weather prediction.

But ill-conceived or untested applications of AI have also caused great harm. We have already seen ways AI systems can amplify, perpetuate, or exacerbate inequitable outcomes. Researchers have shown that AI systems making decisions in high-risk situations, such as credit or housing, can be biased against already disadvantaged communities.

This is why we need to encourage people developing or deploying AI systems to be thoughtful about what they are putting out into the world. We must develop the tools, methodologies, and standards to ensure that AI products and services are safe and secure, accurate, free of harmful bias, and otherwise trustworthy.

Since taking over the gavel of the Research and Technology Subcommittee, I have worked with my colleagues on both sides of the aisle to promote trustworthy AI. I was proud to secure trustworthy AI provisions in the *CHIPS and Science Act*—which the President signed into law last month. My *Promoting Digital Privacy Technologies Act*, which passed the House and awaits a vote in the Senate, supports privacy-enhanced datasets and tools for training AI systems. Additionally, this Committee led the development of the *2020 National AI Initiative Act* to accelerate and coordinate Federal investments in research, standards, and education of trustworthy AI. In that Act, we also directed NIST to develop an AI risk management framework to help organizations understand and mitigate the risks associated with these technologies. I look forward to hearing about the progress of this work and the many other things NIST is doing to promote trustworthy AI in today's discussion.

Academia and industry are also supporting ethical approaches to AI. Universities across the country are adopting principles for responsible use of AI and incorporating ethics into their computer science curricula. Industry is moving past theoretical principles into practical approaches to mitigating AI risks. But there is still much more to do.

I'm looking forward to hearing more about this work from our witnesses today and to discussing what we here in Congress can do to ensure the United States leads the world in trustworthy artificial intelligence. I'd like to again thank our witnesses for joining us today.

Chairwoman STEVENS. With that, the Chair is going to recognize Ranking Member Mr. Feenstra for an opening statement.

Mr. FEENSTRA. Thank you, Chairwoman Stevens, for holding this important hearing today. I very much value of this hearing. And I also want to thank Ranking Member Lucas for attending today. I'm very grateful for that also. And also to the distinguished panel

that we have before us, it's—I appreciate the time and effort that you have taken to come here and to give testimony on this important topic.

Artificial intelligence is fundamentally changing the way we solve some of our society's biggest challenges. From healthcare to transportation, commerce to cybersecurity, AI technologies are revolutionizing almost every aspect of our daily life. But with every new and emerging technology comes new and evolving challenges and risks. Over the years, the Science Committee has held several hearings on AI, discussing challenges ranging from ethics to the work force needs. I hope we can use today's hearing as an opportunity to further these important discussions and shed light on the importance of enabling safe and trustworthy AI.

To do that, we have to first define what makes AI safe and trustworthy, and I believe our witnesses can help us shed light on that today. But in general, I think we can agree that safe and trustworthy AI will meet certain criteria, like including accuracy, privacy, and reliability. Additionally, it is important that trustworthy AI systems utilize robust data, while also protecting the safety and security of the user data.

Some other important factors of trustworthy AI includes transparency, fairness, accountability, and the mitigation of harmful biases. These factors are particularly important to keep in mind as these technologies are being deployed for the use in our daily lives. It is also critical that the data used in AI technologies is accurate because the input data is the foundation, the literal foundation of AI. So that must be our general goal, transparent and fair AI with accurate data and strong privacy protections. We can ensure that by having the standards and evaluation methods in place for these technologies.

The integration of trustworthy AI in key industries has the most potential use and significant competition to advance U.S. industry. AI and other industries of the future like quantum science can revolutionize how business and economics operate, improving efficiency, expanding services, and integrating operations. The key to these benefits, of course, is the trustworthy of AI.

Here in Congress, Members of the Science Committee introduced the bipartisan *National Artificial Intelligence Initiative Act* in 2020, which was made into law through the Fiscal Year 2021 NDAA. The legislation created a broad national security to accelerate investments of responsible AI research, development, and standards, as well as education for AI work force. It facilitated a new public-private partnership to ensure that the United States leads the world in the development and the use of AI systems.

Related to today's hearing, the initiatives require the National Institute of Standards and Technology, NIST, to create the framework for managing risk associated with AI systems and best practices sharing to advance trustworthy AI systems.

As a leader in AI research, measurement, evaluation and standards, NIST has been developing their voluntary AI Risk Management Framework since this last July. The framework has been developed through a consensus-driven, open, transparent, and collaborative process with multiple workshops for industry to provide input. I look forward to hearing more about the progress NIST is

making in implementing this directive and finalizing this important guidance from Ms. Tabassi. I believe that AI risk management from this framework will be critical for our industry to better mitigate risk associated with AI technologies, as well as promote the incorporation of trustworthiness in every stage from design to evaluation of AI technologies.

I'm also looking forward to hearing from the U.S. Chamber of Commerce to learn more about the work through the Commission on the Artificial Intelligence Competitiveness, Inclusion, and Innovation and how they are working to help build customer confidence in AI technologies.

I want to thank our witnesses again for their participation. I thank Madam Chair for putting this hearing on. And with that, I yield back.

[The prepared statement of Mr. Feenstra follows:]

Thank you, Chairwoman Stevens, for holding today's hearing on this important issue.

And thank you, to our distinguished panel of witnesses for joining us heretoday. Artificial intelligence is fundamentally changing the way we solve some of our society's biggest challenges.

From healthcare to transportation; commerce to cybersecurity; A.I. technologies are revolutionizing almost every aspect of daily life. But with every new and emerging technology comes new and evolving challenges and risks. Over the years, the Science Committee has held several hearings on A.I., discussing challenges ranging from ethics to workforce needs.

I hope we can use today's hearing as an opportunity to further these important discussions, and to shed light on the importance of enabling safe and trustworthy A.I. To do that, we have to first define what makes A.I. safe and trustworthy. I believe our witnesses can help shed light on this today.

But in general, I think we can agree that safe and trustworthy A.I. will meet certain criteria like including accuracy, privacy, and reliability. Additionally, it is important that trustworthy A.I. systems utilize robust data while also protecting the safety and security of user data.

Some other important factors of trustworthy A.I. include transparency, fairness, accountability, and mitigation of harmful biases. These factors are particularly important to keep in mind, as these technologies are being deployed for use in our daily lives.

It is also critical that data used by A.I. technologies is accurate because the input data is the foundation of A.I. So that must be our general goal: transparent and fair A.I. with accurate data and strong privacy protections.

We can ensure that by having standards and evaluation methods in place for these technologies. The integration of trustworthy A.I. in key industries has the potential to be a significant competitive advantage for U.S. industry. A.I. and other industries of the future like quantum sciences can revolutionize how businesses and economies operate, improving efficiency, expanding services, and integrating operations. The key to these benefits, of course, is the trustworthiness of A.I.

Here in Congress, Members of the Science Committee introduced the bipartisan *National Artificial Intelligence Initiative Act of 2020*, which was made law through the FY21 NDAA. This legislation created a broad national strategy to accelerate investments in responsible A.I. research, development, and standards, as well as education for the A.I. workforce. It facilitated new public-private partnerships to ensure the U.S. leads the world in the development and use of responsible A.I. systems.

Related to today's hearing, this initiative required the National Institute of Standards and Technology (NIST) to create a framework for managing risks associated with A.I. systems and best practices for sharing data to advance trustworthy A.I. systems. As a leader in A.I. research, measurement, evaluation, and standards, NIST has been developing its voluntary A.I. Risk Management Framework since last July. The framework has been developed through a consensus-driven, open, transparent, and collaborative process with multiple workshops for industry to provide input.

I look forward to hearing more about the progress NIST is making in implementing this directive and finalizing this important guidance from Ms. Tabassi. I believe the A.I. Risk Management Framework will be a critical tool for industry to better mitigate risks associated with A.I. technologies as well as promote the incor-

poration of trustworthiness into every stage from design to evaluation of A.I. technologies.

I am also looking forward to hearing from the U.S. Chamber of Commerce to learn more about their work through the Commission on Artificial Intelligence Competitiveness, Inclusion, and Innovation, and how they are working to help build consumer confidence in A.I. technologies.

I want to thank our witnesses again for their participation. Madam Chair, I yield back.

Chairwoman STEVENS. At some point in time, they will recall and remember that we had today's hearing that is now actually both meeting in person and virtually, so a couple of reminders to Members. First, Members and staff who are attending in person may choose to be masked. It's not a requirement. Any individuals with symptoms, a positive test, or exposure to someone with COVID-19 should wear a mask while present.

Members who are attending virtually should keep their video feed on as long as they're present in the hearing. Members are responsible for their own microphones. Please keep your microphones muted or off unless you are speaking.

Additionally, if Members have documents they wish to submit for the record, please keep them—or please email them to the Committee Clerk, whose email address was circulated prior to the hearing.

If there are Members who wish to submit additional opening statements, your statements will be added to the record at this point.

[The prepared statement of Chairwoman Johnson follows:]

Thank you, Chairwoman Stevens and Ranking Member Feenstra, for holding today's hearing. And welcome to our esteemed panel of witnesses.

We are here today to learn more about the development of trustworthy artificial intelligence and the work being done to reduce the risks posed by AI systems.

Recent advances in computing and software engineering, combined with an increase in the availability of data, have enabled rapid developments in the capabilities of AI systems. These systems are now deployed across every sector of our society and economy, including education, law enforcement, medicine, and transportation. These are sectors for which AI carries the potential for both great benefit, and great harm.

One significant risk across sectors is harmful bias, which can occur when an AI system produces results that are systemically prejudiced. Bias in AI can amplify, perpetuate, and exacerbate existing structural inequalities in our society, or create new ones. The bias may arise from non-representative training data, implicit biases in the humans who design the system, and many other factors. It is often the result of the complex interactions among the human, organizational, and technical factors involved in the development of AI systems. Consequently, the solution to these problems is not a purely technical one. We must ensure that the writing, testing, and deployment of AI systems is an inclusive, thoughtful and accountable process that results in AI that is safe, trustworthy, and free of harmful bias.

That goal remained central in our development of the *National Artificial Intelligence Initiative Act*, which I led alongside Ranking Member Lucas and which we enacted last Congress. In the *National AI Initiative Act*, we directed the National Science Foundation (NSF) to support research and education in trustworthy AI. As we train the next generation of AI researchers, we must not treat ethics as something separate from technology development. The law specifically directs NSF to integrate ethics research and technology education from the earliest stages and establishes faculty fellowships in technology ethics. The recently enacted *CHIPS and Science Act* further directs NSF to require ethics statements in its award proposals to ensure researchers consider the potential societal implications of their work.

As we will learn more about today, the *National AI Initiative Act* also directed the National Institute of Standards and Technology to develop a framework for trustworthy AI, in addition to carrying out measurement research and standards development to enable the implementation of such a framework.

While AI systems continue to make rapid progress, the activities carried out under the *National AI Initiative Act* will be key to grappling with the sociotechnical questions posed by rapidly advancing AI systems.

I look forward to hearing more from our witnesses today and to discussing what more the United States can do to ensure we are the world leader in the development of trustworthy AI. Thank you, and I yield back my time.

Chairwoman STEVENS. And at this time, I'd like to introduce our witnesses. Our first witness is Elham Tabassi. Ms. Tabassi is the Chief of Staff for the Information Technology Laboratory at the National Institute of Standards and Technology. She leads NIST's trustworthy and responsible AI program that aims to cultivate trust in the design, development, and use of AI technologies by improving measurement science, standards, and related tools. Ms. Tabassi is a member of the National AI Research Task Force and has been at NIST since 1999.

Our next witness is Dr. Charles Isbell. Dr. Isbell is the Dean and John P. Imlay, Jr. Chair of the College of Computing at Georgia Tech. His recent work focuses on building autonomous systems that can interact with large numbers of other intelligence agents, including humans and AI systems. Dr. Isbell also studies the effects of AI bias and pursues reform in computing education, focusing on broadening participation and access. He is an elected fellow of AAAI (Association for the Advancement of Artificial Intelligence), ACM (Association for Computing Machinery), and the American Academy of Arts and Sciences.

Our third witness is Mr. Jordan Crenshaw. Mr. Crenshaw serves as the Vice President of the U.S. Chamber of Commerce's Technology Engagement Center. He also manages the Chamber's Privacy Working Group and which is comprised of nearly 300 companies and trade associations in which developed model privacy legislation and principles. Prior to his current position, Mr. Crenshaw led the Chamber's Telecommunication and E-Commerce Policy Committee, which analyzes Federal privacy, cloud computing, broadband internet, e-commerce and broadcast policies.

Our final witness is Ms. Navrina Singh. Ms. Singh is the Founder and Chief Executive Officer (CEO) of Credo. Credo AI helps organizations to monitor, measure, and manage AI introduce risk. Prior to co-founding Credo AI, Ms. Singh was the Director and Principal of Product in Microsoft Cloud and AI, where she built natural language-based conversational AI products. Currently, Ms. Singh serves as a member of the National AI Advisory Committee, which is tasked with advising the President and the National AI Initiative Office on topics related to the National AI Initiative.

As our witnesses know—should know, you will each have 5 minutes for your spoken testimony. Your written testimony will be included in the record for the hearing. They're great testimonies. When you have completed your spoken testimony, we'll begin with questions. Each Member will have 5 minutes to question the panel.

We will start with Ms. Tabassi.

**TESTIMONY OF MS. ELHAM TABASSI, CHIEF OF STAFF,
INFORMATION TECHNOLOGY LABORATORY,
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY**

Ms. TABASSI. Good morning, Chairwoman Stevens, Ranking Member Feenstra, and distinguished Members of the Sub-

committee. I am Elham Tabassi, and I serve as the lead for the Trustworthy and Responsible AI program at the Department of Commerce's National Institute of Standards and Technology known as NIST. Thank you for the opportunity to testify today on NIST's effort to advance the trustworthy and responsible development and use of artificial intelligence. This Committee is well aware of the importance of advancing research and standards to cultivate trust in AI. Thank you for your dedication to this important issue and for your support of NIST's role.

Artificial Intelligence holds the promise to revolutionize and enhance our society and economy, but the development and use of these systems are not without challenges or risks. Through robust collaboration with stakeholders across government, industry, civil groups, and academia, NIST works to advance research, standards, measurements, and tools to manage these risks and realize the full promise of this technology for all Americans.

Among its work, NIST is developing the AI Risk Management Framework, or AI RMF, to provide guidance on mapping, measuring, and managing risks associated with AI. Like the well-known cybersecurity and privacy frameworks, the AI RMF will provide a set of outcomes that enable dialog, understanding, and actions to manage AI risks. Critically, the framework will focus on managing risks not just to organizations, but also to individuals and society. This approach is reflective of the sociotechnical nature of AI systems as a product of the complex human, organizational, and technical factors involved in their design and development.

As is the case with all our publications, NIST is taking a stakeholder-driven and open process to coordinate the development of the framework. From the start of this initiative last year, NIST has engaged a broad range of stakeholders, including through several workshops and public comment opportunities. Based on stakeholder feedback, and consistent with congressional direction, NIST is on track to publish the final AI RMF 1.0 in January 2023. The technology and standards landscape for AI will continue to evolve. Therefore, NIST intends for the framework and related guidance to be updated over time to reflect new knowledge, awareness, and practices.

Building off the RMF there is much more work to do to develop additional guidance, standards, measures, and tools to evaluate and measure AI trustworthiness, especially for specific characteristics and use cases. For example, NIST has significantly expanded its research efforts to mitigate harmful bias with a focus on sociotechnical approach.

To support the advancement of AI standards, NIST seeks to bolster knowledge, leadership, and coordination on AI, including by engaging with other government agencies within United States and internationally. NIST engages with partners around the world, including through the Organization for Economic Cooperation and Development, OECD, and the U.S.-EU Trade and Technology Council (TTC) to advance shared goals in trustworthy and responsible AI.

NIST also coordinates with other Federal agencies and leads several policymaking and interagency efforts. This includes administering the National Artificial Intelligence Advisory Committee or

NAIAC, which advises the President and the National AI Initiative Office.

Advancing research and standards that contribute to more secure, private, fair, rights-affirming, and world-leading digital economy is a top priority for NIST. Thank you for the opportunity to present on NIST's activities to improve trustworthy and responsible AI. I look forward to your questions.

[The prepared statement of Ms. Tabassi follows:]

Testimony of

Elham Tabassi

Chief of Staff
Information Technology Laboratory

National Institute of Standards and Technology
United States Department of Commerce

Before the
United States House of Representatives
Committee on Science, Space, and Technology
Subcommittee on Research and Technology

Trustworthy AI: Managing the Risks of Artificial Intelligence

September 29, 2022

Chairwoman Stevens, Ranking Member Feenstra, and distinguished members of the Subcommittee, I am Elham Tabassi, Chief of Staff of the Information Technology Laboratory (ITL) and the lead for NIST's trustworthy and responsible AI program at the Department of Commerce's National Institute of Standards and Technology – known as NIST. We appreciate the committee's continued support of our work and thank you for the opportunity to testify today on NIST's efforts to improve the trustworthiness of artificial intelligence.

NIST is home to five Nobel Prize winners, with programs focused on national priorities such as cybersecurity, advanced manufacturing, semiconductors, the digital economy, precision metrology, quantum information science, biosciences and artificial intelligence. NIST's mission is to promote U.S. innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve our quality of life.

In the NIST Information Technology Laboratory, we work to cultivate trust in information technology and metrology. Trust in the digital economy is built upon key attributes like cybersecurity, privacy, usability, interoperability, equity, and avoiding bias and increasing usefulness in the development and deployment of technology. NIST conducts fundamental and applied research, advances standards to understand and measure limits and capabilities of technology and develops tools to evaluate such measurements. Technology standards and measurements—and the foundational and applied research that enables their development and use—are critical to advancing trust in digital products and services. These standards and measurements can provide increased assurance and utility, thus enabling more secure, private, and rights-affirming technologies.

NIST's Role in Artificial Intelligence

NIST contributes to the research, standards, measurements, and data required to realize the full promise of artificial intelligence (AI) as a tool that will enable American innovation, enhance economic security, and improve our quality of life.

As a non-regulatory agency, NIST prides itself on the strong partnerships it has cultivated with the government and private sector. NIST seeks and relies on diverse stakeholder feedback among government, industry, academia, and non-profit entities to develop and improve its resources. The collaborative, transparent, and open processes NIST uses to develop resources result in more effective and usable resources that are trusted, and therefore, widely used by various organizations. Our resources are used by federal agencies, as well as private sector organizations of all sizes, educational institutions, and state, local, tribal, and territorial governments.

Much of NIST's AI effort¹ focuses on cultivating trust in the design, development, and use of AI technologies and systems. Working with the community, NIST is:

- conducting fundamental research to advance trustworthy AI technologies and understand and measure their capabilities and limitations
- applying AI research and innovation across NIST laboratory programs
- establishing benchmarks and developing data and metrics to evaluate AI technologies
- leading and participating in the development of technical AI standards

¹ <https://www.nist.gov/artificial-intelligence>

- contributing to discussions and development of AI policies, including supporting the National AI Advisory Committee²

NIST AI Risk Management Framework

Among its many AI-related activities, NIST is developing the AI Risk Management Framework³ (AI RMF) to provide guidance on managing risks to individuals, organizations, and society associated with AI. AI risk management is about offering a path to minimize potential negative impacts of AI systems, as well as pointing to opportunities to maximize positive impacts and creating opportunities for innovation. Identifying, mitigating, and minimizing risks and potential harms associated with AI technologies are essential steps towards the development of trustworthy AI systems and their appropriate and responsible use. Like NIST's well-known Cybersecurity and Privacy Frameworks, the NIST AI RMF will provide a set of outcomes that enable dialogue, understanding, and actions to manage AI risks. The AI RMF is a voluntary framework seeking to provide a flexible, structured, and measurable process to address AI risks prospectively and continuously throughout the AI lifecycle.

In August, NIST released its second draft of the AI RMF⁴ with the goal of releasing AI RMF 1.0 in January. This is consistent with congressional direction in the National Artificial Intelligence Act of 2020. This latest draft builds on the March 2022 initial draft and a December 2021 concept paper – and the many comments from organizations and individuals.

NIST also released a draft AI RMF Playbook⁵ in August. This companion to the AI RMF when completed will provide additional guidance to organizations on the actions they can take to meet the outcomes included in the Framework.

AI research and development, as well as the standards landscape, are evolving rapidly. For that reason, the AI RMF and its related documents will evolve over time and reflect new knowledge, awareness, and practices. NIST intends to continue its robust engagement with stakeholders to keep the Framework up to date with AI trends and reflect experience based on the use of the AI RMF. Ultimately, the AI RMF will be offered in multiple formats, including online versions, to provide maximum flexibility.

The Framework is being developed through a consensus-driven, open, transparent, and collaborative process. From the start of this initiative, NIST has offered a broad range of stakeholders the opportunity to take part in workshops⁶, respond to a Request for Information (RFI)⁷, and review draft reports⁸ and other documents including draft approaches⁹ and versions of the framework¹⁰. NIST also has reached out directly to AI practitioners along with other stakeholders across a full spectrum of interests domestically and internationally. This outreach

² <https://www.nist.gov/artificial-intelligence/national-artificial-intelligence-advisory-committee-naiac>

³ <https://www.nist.gov/itl/ai-risk-management-framework>

⁴ https://www.nist.gov/system/files/documents/2022/08/18/AI_RM_F_2nd_draft.pdf

⁵ <https://pages.nist.gov/AIRMF/>

⁶ <https://www.nist.gov/itl/ai-risk-management-framework/ai-risk-management-framework-workshops-events>

⁷ <https://www.nist.gov/itl/ai-risk-management-framework/ai-rmf-development-request-information>

⁸ <https://www.nist.gov/system/files/documents/2022/03/17/AI-RMF-1stdraft.pdf>

⁹

https://www.nist.gov/system/files/documents/2021/12/14/AI%20RMF%20Concept%20Paper_13Dec2021_posted.pdf

¹⁰ <https://www.nist.gov/itl/ai-risk-management-framework>

has included companies, government agencies, academia, and not-for-profit organizations representing civil society, consumers, and industry. NIST has actively encouraged others to provide direct input, and many organizations and individuals have contributed their insights to NIST. Those have included international organizations, with the goal of aligning the NIST Framework with standards and approaches being developed around the globe.

The current draft AI RMF defines certain key characteristics of trustworthy AI systems and offers guidance for mapping, measuring, and managing them. As defined in the draft AI RMF, trustworthy AI is valid and reliable, safe, fair, and bias is managed, secure and resilient, accountable and transparent, explainable and interpretable, and privacy-enhanced. AI systems are socio-technical in nature, meaning they are a product of the complex human, organizational, and technical factors involved in their design, development, and use. Many of the trustworthy AI characteristics – such as bias, fairness, interpretability, and privacy – are directly connected to societal dynamics and human behavior.

NIST’s Research on AI Trustworthiness Characteristics

To build on NIST’s work on the AI RMF and provide additional guidance to organizations to advance trustworthy and responsible AI, NIST also conducts fundamental research on many of the AI trustworthiness characteristics.

» *AI Trustworthiness Characteristics – Fair and Bias is Managed*

While there are many approaches for ensuring technologies that we use every day are safe and secure, there is less research into how to advance systems that are fair with bias managed. Fairness in AI includes concerns for equality and equity by addressing issues such as bias and discrimination. Standards of fairness can be complex and difficult to define because perceptions of fairness differ among cultures and may shift depending on application and context of use.

NIST has significantly expanded its research efforts to identify, understand, measure, manage and mitigate bias, with a focus on a socio-technical approach. NIST recently published “Towards a Standard for Identifying and Managing Bias in Artificial Intelligence” (NIST Special Publication 1270)¹¹, which identifies the concepts and challenges associated with bias in AI and provides preliminary guidance for addressing them.

NIST has identified three major categories of AI bias to be considered and managed: systemic, computational, and human, all of which can occur in the absence of prejudice, partiality, or discriminatory intent. Current attempts for addressing the harmful effects of AI bias remain focused largely on computational factors such as representativeness of datasets and fairness of machine learning algorithms. Human and systemic institutional and societal factors are significant sources of AI bias that are currently overlooked. Systemic bias can be present in AI datasets, the organizational norms, practices, and processes across the AI lifecycle, and the broader society that uses AI systems. Human biases relate to how an individual or group perceives and uses AI system information to make a decision or fill in missing information.

Through the NIST National Cybersecurity Center of Excellence (NCCoE), we are beginning a project, “Mitigation of AI/ML Bias in Context”¹², to develop additional guidance to mitigate bias in AI and Machine Learning (ML). Under the NCCoE model, NIST works collaboratively with

¹¹ <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf>

¹² <https://www.nccoe.nist.gov/projects/mitigating-aiml-bias-context>

relevant industry and academia partners. The “Mitigation of AI/ML Bias in Context,” project intends to apply the concepts in our March 2022 NIST publication on bias to build a proof-of-concept implementation, or “use case,” for credit underwriting decisions in the financial services sector. Future application use cases may also be considered, such as hiring or school admissions. These will help promote fair and positive outcomes that benefit users of AI/ML services, the organizations that deploy them, and all of society. A small but novel part of this project will examine the interplay between bias and cybersecurity, with the goal of identifying approaches which might mitigate risks that exist across these two critical characteristics of trustworthy AI.

» *AI Trustworthiness Characteristics – Explainable and Interpretable*

Explainability and interpretability are important characteristics to ensure users and operators of AI can understand the decisions or predications made by AI, thus avoiding the “opaque system” concept associated with AI. Explainability refers to a representation of the mechanisms underlying an algorithm’s operation, whereas interpretability refers to the meaning of an AI systems’ output in the context of its designed functional purpose.

NIST has released two publications aimed at providing deeper understanding of the principles of Explainability and interpretability: “Four Principles of Explainable Artificial Intelligence” (NISTIR 8312)¹³ and “Psychological Foundations of Explainability and Interpretability in Artificial Intelligence” (NISTIR 8367)¹⁴.

» *AI Trustworthiness Characteristics – Secure and Resilient*

AI systems that can withstand adversarial attacks and maintain confidentiality, integrity, and availability are resilient and secure systems.

NIST released the draft “A Taxonomy and Terminology of Adversarial Machine Learning”¹⁵ (NISTIR 8269) to advance a taxonomy for securing applications of AI, specifically, adversarial machine learning. NIST’s Cybersecurity Framework¹⁶ is widely used to address the cybersecurity risks of organizations. NIST is constantly updating the Cybersecurity Framework to account for changes in the cybersecurity technology, standards, and risk landscape.

NIST is building an experimentation testbed called Dioptra¹⁷ to begin to evaluate adversarial attacks against ML algorithms. The testbed aims to facilitate security evaluations of ML algorithms under a diverse set of conditions. To that end, the testbed has a modular design enabling researchers to easily swap in alternative datasets, models, attacks, and defenses. The result is the ability to advance the metrology needed to ultimately help secure AI systems.

» *AI Trustworthiness Characteristics – Privacy-enhanced*

Privacy safeguards the important human values of autonomy and dignity through methods that focus on providing individuals with anonymity, confidentiality, and control over various facets of their identities. These outcomes generally should guide choices for AI system design,

¹³ <https://www.nist.gov/publications/four-principles-explainable-artificial-intelligence>

¹⁴ <https://www.nist.gov/publications/psychological-foundations-explainability-and-interpretability-artificial-intelligence>

¹⁵ <https://www.nccoe.nist.gov/ai/adversarial-machine-learning>

¹⁶ <https://www.nist.gov/cyberframework>

¹⁷ <https://pages.nist.gov/dioptra/>

development, and deployment. From a policy perspective, privacy-related risks may overlap with security, bias, and transparency.

NIST's Privacy Risk Assessment Methodology¹⁸, developed in 2016 and NIST's Privacy Framework¹⁹, issued in 2020, are voluntary tools that organizations from all industry sectors across the world are using to identify and manage privacy risks in the systems, products and services they develop and deploy, improve their privacy programs, and better comply with privacy regulation.

NIST is also conducting research on privacy-enhancing technologies (PETs) to advance data-driven, innovative solutions to preserve the right to privacy, including hosting the Privacy Engineering Collaboration Space²⁰, a virtual public platform that serves as a clearinghouse for open-source tools and PETs use cases. In coordination with the National Science Foundation (NSF) and the White House Office of Science and Technology Policy (OSTP), NIST is co-sponsoring the U.S.-U.K. prize competition on PETs²¹. First announced at the Summit for Democracy in December 2021, the winning solutions will compete for a combined U.S.-U.K. prize pool of \$1.6 million and will be showcased at the second Summit for Democracy anticipated in early 2023.

Research on Applications of AI

NIST's multidisciplinary laboratories and varied fields are an ideal environment to develop and apply AI²². Various AI techniques are being used to support NIST scientists and engineers, drawing on ML and AI tools to gain a deeper understanding of and insight into our research. NIST is integrating AI into the design, planning, and optimization of NIST's research efforts – including hardware for AI²³, computer vision, engineering biology and biomanufacturing, image and video understanding, medical imaging, materials science, manufacturing, disaster resilience, energy efficiency, natural language processing, biometrics, quantum science, robotics, and advanced communications technologies. Key focus areas include innovative measurements using AI/ML techniques, predictive systems using AI/ML models, and enabling and reducing the barriers to autonomous measurement platforms.

AI Measurement and Evaluation

NIST has a long history of devising appropriate metrics, measurement tools, and challenge problems to support technology development. NIST first started the measurement and evaluation of automated fingerprint identification systems in the 1960s. Evaluations strengthen research communities, establish research methodology, support the development of standards, and facilitate technology transfer. NIST is looking to bring these benefits of community evaluations to bear on the problem of constructing trustworthy AI systems. These evaluations will begin with community input to identify potential harms of selected AI technologies in context, and the data

¹⁸ <https://www.nist.gov/itl/applied-cybersecurity/privacy-engineering/resources>

¹⁹ <https://www.nist.gov/privacy-framework/privacy-framework>

²⁰ <https://www.nist.gov/itl/applied-cybersecurity/privacy-engineering/collaboration-space>

²¹ <https://www.nist.gov/itl/applied-cybersecurity/privacy-engineering/collaboration-space/prize-challenges>

²² <https://www.nist.gov/applied-ai>

²³ <https://www.nist.gov/artificial-intelligence/hardware-ai>

requirements for AI evaluations. NIST also hosts a biweekly AI metrology colloquia series²⁴, where leading researchers share current work on AI measurement and evaluation.

As discussed above, NIST has been engaged in focused efforts to establish common terminologies, definitions, and taxonomies of concepts pertaining to characteristics of AI technologies in order to form the necessary underpinnings for trustworthy AI systems. Each of these characteristics also requires its own portfolio of measurements and evaluations. For each characteristic, NIST aims to document and improve the definitions, applications, and strengths and limitations of metrics and measurement methods in use or being proposed. NIST's current efforts represent only a small portion of the research that will be required to test and evaluate trustworthy AI systems.

A significant challenge in the evaluation of trustworthy AI systems is that context (the specific use case) matters; accuracy measures alone will not provide enough information to determine if deploying a system is warranted. The accuracy measures must be balanced by the associated risks or societal harms that could occur. The tolerance for error drops as the potential impacts of risk rise.

New NIST efforts in AI evaluation will focus on other socio-technical aspects of system performance in addition to accuracy. In particular, the evaluations have the goal of identifying risks and harms of systems before such systems are deployed, and to define (and eventually create) data sets and evaluation infrastructure that will allow system builders to detect the extent to which their system exhibits those harms.

Examples of NIST AI measurement and evaluation projects²⁵ include:

- *Biometrics*: Over that past sixty years, NIST has been testing and evaluating biometric recognition technologies, including face recognition, fingerprint, biometric quality, iris recognition, and speaker recognition.
- *Computer vision*: NIST's computer vision program includes several activities contributing to the development of technologies that extract information from image and video streams through systematic, targeted annual evaluations and metrology advances, including the Open Medica Forensics Challenge, Activities in Extended Video (ActEV), handwriting recognition and translation evaluation, and others.
- *Information retrieval*: The information retrieval research uses large, human-generated text, speech, and video files to create test collections through the Text Retrieval (TREC), TREC Video Retrieval Evaluation (TRECVID), and Text Analysis (TAC) Conferences. The Text Retrieval Conference is responsible for significant advancements in search technology. A 2010 NIST study²⁶ estimated that without TREC, U.S. internet users would have spent an estimated 3.5 billion worth of additional hours using search engines between 1999 and 2009.

AI Standards

NIST plays a critical role in the standards process as the nation's measurement laboratory and has a unique role relating to standards in the Federal enterprise. Our coordination function,

²⁴ <https://www.nist.gov/programs-projects/ai-measurement-and-evaluation/ai-metrology-colloquia-series>

²⁵ <https://www.nist.gov/programs-projects/ai-measurement-and-evaluation/nist-ai-measurement-and-evaluation-projects>

²⁶ <https://trec.nist.gov/pubs/2010.economic.impact.pdf>

currently defined under the National Technology Transfer and Advancement Act and the NIST Organic Act, has yielded benefits to the nation ever since the Institute was established by Congress as the National Bureau of Standards in 1901. NIST's strong ties to industry and the standards development community have enabled NIST to take on critical standards-related challenges and deliver timely and effective solutions.

NIST works to support the development of AI standards that promote innovation and public trust in systems that use AI. Pursuant to U.S. Leadership in AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools²⁷, NIST seeks to bolster AI standards-related knowledge, leadership, and coordination; conduct research to support development of technically sound standards for trustworthy AI; promote partnerships to develop and use standards; and engage internationally to advance AI standards.

I serve as the Federal AI Standards Coordinator to work across the government and industry stakeholders to gather and share information on AI standards-related needs, strategies, and best practices.

NIST facilitates federal agency coordination in the development and use of AI standards in part through the Interagency Committee on Standards Policy (ICSP) AI Standards Coordination Working Group²⁸. This working group seeks to foster agency interest and participation in AI standards and conformity assessment activities, facilitate coordination of U.S. government positions on draft standards, identify effective means of coordinating with and contributing towards voluntary consensus bodies, align U.S. government activities with those of the private sector on AI standards development activities, promote effective and consistent federal policies leveraging AI standards, and raise awareness of federal agencies' use of AI that contributes to standards activities.

NIST also engages internationally through bilateral and multilateral work on AI. The United States championed development of the first international principles for the responsible use of AI at the Organisation for Economic Co-operation and Development, or OECD. The U.S. also serves as a founding member of the Global Partnership on AI, which includes all members of the G7 and others such as Brazil and India, to coordinate R&D AI initiatives. NIST advances research on trustworthy AI with the Indo-Pacific Economic Framework. NIST supports the US-EU Trade and Technology Council (TTC) in building common approaches for trustworthy AI. Under the TTC, the U.S. and EU have launched a new AI sub-working group where NIST is working towards common frameworks for AI risk management and developing metrics and methodologies for measuring AI trustworthiness. And as mentioned above, the U.S. – led by NIST, NSF, and OSTP – is collaborating with the UK to develop prize challenges on advancing privacy-enhancing technologies.

Interagency Coordination

NIST leads and participates in several federal AI policymaking efforts and engages with many other federal offices and interagency groups. This includes administering the National Artificial Intelligence Advisory Committee (NAIAC)²⁹, on behalf of the Department of Commerce. The NAIAC is tasked with advising the President and the National AI Initiative Office. NIST supports the operation of this advisory committee. The Secretary of Commerce appointed the 27

²⁷ https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf

²⁸ <https://www.nist.gov/standardsgov/icsp-ai-standards-coordination-working-group-aiscwg-charter>

²⁹ <https://www.nist.gov/artificial-intelligence/national-artificial-intelligence-advisory-committee-naiac>

members in April 2022. NAIAC held its first meeting in May 2022. Five working groups have been established to focus NAIAC's work on leadership in trustworthy AI, leadership in research and development, supporting the U.S. workforce and providing opportunity, U.S. leadership and competitiveness, and international cooperation.

NIST also co-chairs the National Science and Technology Council's Machine Learning and Artificial Intelligence Subcommittee³⁰, the Networking and Information Technology Research and Development's (NITRD) AI Working group³¹, and the NITRD Fast Track Action Committee³² which is drafting a national strategy to advance privacy-preserving data sharing and analytics. NIST founded and is co-chairing the AI Standards Coordination Working Group (AISCWG) under the Interagency Committee on Standards Policy (ICSP). NIST's AI lead also serves as Federal AI Standards Coordinator and is a member of the National AI Research Resource Task Force³³.

Conclusion

Advancing artificial intelligence research and standards that contribute to a secure, private, interoperable, and world-leading digital economy is a top priority for NIST. Our economy is increasingly global, complex, and interconnected. It is characterized by rapid advances in technology. The timely availability of AI trustworthiness standards and guidance is a dynamic and critical challenge. Through robust collaboration with stakeholders across government, industry, and academia in the U.S. and elsewhere, NIST aims to cultivate trust and foster an environment that enables AI innovation on a global scale – and to do so in a way that respects and advances human rights.

NIST's team includes some of the top AI and standards experts in the world. This includes staff with multidisciplinary backgrounds in science and engineering. Working with our partners in other federal agencies, the private sector, academia, and other allied countries, and with the support of Congress, we will work tirelessly to address current and future challenges.

Thank you for the opportunity to present on NIST activities to improve AI trustworthiness. I look forward to your questions.

³⁰ https://www.ai.gov/about/#MLAI-SC_Machine_Learning_and_AI_Subcommittee

³¹ <https://www.ai.gov/a-new-nitrd-iwg-for-artificial-intelligence-ai-rd/>

³² <https://www.nitrd.gov/coordination-areas/privacy-rd/appdsa/>

³³ <https://www.ai.gov/naiac/>



Elham Tabassi (Fed)
Chief of Staff, Information Technology Laboratory

Elham Tabassi is the Chief of Staff in the Information Technology Laboratory (ITL) at the National Institute of Standards and Technology (NIST). She leads NIST Trustworthy and Responsible AI program that aims to cultivate trust in the design, development, and use of AI technologies by improving measurement science, standards, and related tools in ways that enhance economic security and improve quality of life. She has been working on various machine learning and computer vision research projects with applications in biometrics evaluation and standards since she joined NIST in 1999. She is the principal architect of NIST Fingerprint Image Quality (NFIQ) which is now an international standard for measuring fingerprint image

quality and has been deployed in many large scale biometric applications worldwide. She is a member of the National AI Resource Research Task Force, a senior member of IEEE, and a fellow of Washington Academy of Sciences.

PUBLICATIONS

[NIST Fingerprint Image Quality 2](#)

JULY 13, 2021

AUTHOR(S)

ELHAM TABASSI, MARTIN OLSEN, OLIVER BAUSINGER, CHRISTOPH BUSCH, ANDREW FIGLARZ, **GREGORY FIUMARA**, OLAF HENNIGER, JOHANNES MERKLE, TIMO RUHLAND, CHRISTOPHER SCHIEL, MICHAEL SCHWAIGER

NIST Fingerprint Image Quality (NFIQ 2) is open source software that links image quality of optical and ink 500 pixel per inch fingerprints to operational

[NIST Special Database 302: Nail to Nail Fingerprint Challenge](#)

DECEMBER 11, 2019

AUTHOR(S)

GREGORY P. FIUMARA, PATRICIA A. FLANAGAN, JOHN D. GRANTHAM, KENNETH KO, KAREN MARSHALL, MATTHEW SCHWARZ, **ELHAM TABASSI**, BRYAN WOODGATE, CHRISTOPHER BOEHNEN

In September 2017, the Intelligence Advanced Research Projects Activity (IARPA) held a data collection as part of its Nail to Nail (N2N) Fingerprint Challenge

[Nail to Nail Fingerprint Challenge: Enrollment Set Size Variability](#)

JUNE 24, 2019

AUTHOR(S)

GREGORY P. FIUMARA, KENNETH KO, **ELHAM TABASSI**, PATRICIA A. FLANAGAN, JOHN D. GRANTHAM, KAREN MARSHALL, MATTHEW SCHWARZ, **BRYAN WOODGATE**

In September 2017, the Intelligence Advanced Research Projects Activity held a fingerprint data collection as part of the Nail to Nail Fingerprint Challenge

NIST Special Database 301: Nail to Nail Fingerprint Challenge Dry Run

JULY 11, 2018

AUTHOR(S)

GREGORY P. FIUMARA, PATRICIA A. FLANAGAN, MATTHEW SCHWARZ, ELHAM TABASSI, CHRISTOPHER BOEHNEN

In April 2017, the Intelligence Advanced Research Projects Activity (IARPA) held a dry run for the data collection portion of its Nail to Nail (N2N) Fingerprint

Nail to Nail Fingerprint Challenge: Prize Analysis

MAY 3, 2018

AUTHOR(S)

GREGORY P. FIUMARA, ELHAM TABASSI, PATRICIA A. FLANAGAN, JOHN D. GRANTHAM, KENNETH KO, KAREN MARSHALL, MATTHEW SCHWARZ, BRYAN WOODGATE, CHRISTOPHER BOEHNEN

In September 2017, the Intelligence Advanced Research Projects Activity held a fingerprint data collection as part of the Nail to Nail Fingerprint Challenge

Chairwoman STEVENS. Dr. Isbell.

**TESTIMONY OF DR. CHARLES ISBELL,
DEAN AND JOHN P. IMLAY, JR. CHAIR
OF THE COLLEGE OF COMPUTING,
GEORGIA INSTITUTE OF TECHNOLOGY**

Dr. ISBELL. Thank you, Subcommittee Chair Stevens, Ranking Members Feenstra and Lucas, and distinguished Members of the Subcommittee. I'm Charles Isbell. I'm a Professor in and Dean for the College of Computing at Georgia Tech. Thank you for the opportunity to be here today.

So by way of explaining my background, let me note that while I tend to focus on statistical machine learning, my research passion is actually interactive artificial intelligence. As noted at the top of the hearing, there, the fundamental research goal is to understand how to build autonomous agents who must live and interact with large numbers of other intelligent agents, some of whom may be human. But I'm also an educator. As such, I spend much of my energy focusing on providing access to all those who wish to be a part of this ongoing conversation around the role of AI and computing in our lives. My discussion today and answers to your questions you ask will be informed by both my research and educator selves.

So let us begin this discussion by defining our terms. There are many potential definitions of AI. My favorite one is that it is the art and science of making computers act the way they do in the movies. In the movies, computers are often semi-magical and anthropomorphic. They do things that if humans did them, we would say they required intelligence.

This definition is borne out in our use of AI in the everyday world. We use the infrastructure of AI to search billions upon billions of documents to find the answers to a staggering variety of questions, often expressed literally as questions. We use automatically tagged images to organize our photos. And we use that same infrastructure to plan optimal routes for trips, even altering our routes on the fly in the face of changes in traffic. In fact, we let our cars mostly drive themselves in that very same traffic playing the role of a tireless chauffeur.

As noted by the Chair, we're able to automatically detect tumors from X-rays, even those that are trained—that trained doctors find difficult to see. We let computers finish our sentences as we type text and use search engines, sometimes facilitating a subtle shift from prediction of our behavior to influence over our behavior. Often, we take advantage of these services by using our phones to interpret a wide variety of spoken commands.

So in some very important sense, AI already exists. It is not the AI of fanciful science fiction, neither benevolent intelligence working with humans as we traverse the galaxy, nor malevolent AI that seeks humanity's destruction. Nonetheless, we are living every day with machines who make decisions that if humans made them, we would attribute to intelligence. And the machines often make those decisions faster, and some might argue better, than humans would.

Yet like all computing systems, at bottom, AI simply makes us more efficient. It amplifies our ability to make decisions, including bad ones, all too often automating the biases baked into our data

and that of its developers. By way of example, according to the Marshall Project, most States use some form of automated risk assessment at some stage in the criminal justice system. We set out to predict recidivism as if that means the chance of committing a crime again, when in fact, what we're actually predicting is the chance of being arrested and convicted again. As with the shift from predicting behavior to influencing it, this distinction is subtle, but important. Without recognition of the difference, one can create a feedback loop and make things worse, without even noticing it.

Although we sometimes act as if the machine is doing the work, it is worth noting that these machines are making decisions with us, with humans. They are partners, and as with any partner, it is important that we understand what our partner is doing and why. To make AI trustworthy, we need a more informed citizenry, something we can accomplish by requiring that our AI partners are more transparent on the one hand, but that we are more savvy on the other.

So speaking of definitions, by transparency, I mean that an AI algorithm should be inspectable, that the kind of data the algorithm uses to build its model should be available, and the decisions that such algorithms make should be understandable. In other words, as we deploy these algorithms, each algorithm should be able to explain its output. "This applicant was assigned this score because" is more useful and less prone to misuse than just "This applicant was assigned this score."

But to really understand such machines, much less to create them, we should strive for all of our citizens to not only be literate, but to be competent. That is, they must understand computing and computational thinking and how it fits into problem solving in their everyday lives. In the long term, one of the key solutions to AI bias will be bringing a wider group of people into computing education and into machine learning more specifically. We have to improve the number and the diversity of those entering the field and participating in and influencing the conversation because it is the right thing to do, but also because it is the only way for us to compete.

It should not be lost that putting these two thoughts together suggests that the process by which we build AI algorithms is a shared effort that requires a wide swath of citizens to be informed and engaged and for developers to accept the responsibility for including the users of and sometimes targets of those systems in the development process itself. As a field, we have not caught up to the reality of the responsibility that we hold, and it is something that we simply must do. We must move from tool sets and skill sets to mindsets, incorporating responsibility in all that we do from the ground up.

I'm very excited for this hearing. I think advances in AI are essential to our economic and social future. These are all areas in which funding—the funding power of the National Science Foundation and NIST as well can make a huge difference. So thank you very much, and I look forward to your questions.

[The prepared statement of Dr. Isbell follows:]

Subcommittee Chair Stevens, Subcommittee Ranking Member Feenstra, Committee Chair Johnson, Ranking Member Lucas, and distinguished members of the subcommittee, my name is Dr. Charles Isbell and I am a Professor in and Dean for the College of Computing at Georgia Tech. Thank you for the opportunity to appear before this Subcommittee to discuss:

1. The importance of a culture of responsibility around artificial intelligence (AI) systems.
2. The need for transparency in AI systems in order to identify harmful bias.
3. Mitigation of the risks in AI.

By way of explaining my background, let me note that while I tend to focus on statistical machine learning, my research passion is actually artificial intelligence. I like to build large integrated systems, so I also tend to spend a great deal of my time doing research on autonomous agents, interactive entertainment, some aspects of human-computer interaction, software engineering, and even programming languages

I think of my field as interactive artificial intelligence. My fundamental research goal is to understand how to build autonomous agents that must live and interact with large numbers of other intelligent agents, some of whom may be human. Progress towards this goal means that we can build artificial systems that work with humans to accomplish tasks more effectively; can respond more robustly to changes in environment, relationships, and goals; and can better co-exist with humans as long-lived partners.

As the members of this Subcommittee well know, there has been an explosion in the development and deployment of what we might call AI technology. With that explosion has come a corresponding explosion in interest in AI.

In any discussion—particularly technical ones—it helps to define our terms. There are many potential definitions of AI. My favorite one is that it is “the art and science of making computers act like they do in the movies.” In the movies, computers are often semi-magical and anthropomorphic; they do things that, if humans did them, we would say they required intelligence.

This definition is borne out in our use of AI in the everyday world. We use the infrastructure of AI to search billions upon billions of documents to find the answers to a staggering variety of questions—often expressed literally as questions. We use automatically tagged images to organize our photos, and we use that same infrastructure to plan optimal routes for trips—even altering our routes on-the-fly in the face of changes in traffic. We are able to automatically detect tumors from x-rays, even those that trained doctors find difficult to see. We let computers finish our sentences as we type texts and use search engines, sometimes facilitating a subtle shift from prediction of our behavior to influence over our behavior. Often we take advantage of these services by using our phones (our phones!) to interpret a wide variety of spoken commands.

So, in some very important sense, AI already exists. It is not the AI of science fiction, neither benevolent intelligences working with humans as we traverse the galaxy, nor malevolent AI that seeks humanity's destruction. Nonetheless, we are living every day with machines that make decisions that, if humans made them, we would attribute to intelligence. And the machines often make those decisions faster and better than humans would.

Importantly, each of the examples we consider above is a distinctly human-centered problem. It is human-centered both in the sense that these systems are trying to solve problems that humans deal with every day—question answering, symptom evaluation, navigation—but also human-centered in the sense that humans have or currently perform some of those tasks. Presumably, these developments are all to the good. We are living up to the promise of technology that allows us to automate away work that is dirty, dangerous, or dull, freeing up human capital to be more productive, and, hopefully, for humans to be more fulfilled. The social and economic benefits are potentially immense.

There are also some reasons for concern. Those who work in the field will tell you that very often they aren't sure exactly how their algorithms reach the correct answer, only that they do. AI scientists describe these algorithms as "black box models."

The second concern is that sometimes those algorithms reach the wrong conclusion, and in a way that harms people and society. Artificial intelligence has all too often automated the biases of its programmers, or baked into its data. As a result, AI products have already been caught making biased decisions in banking, hiring, health care and criminal justice.

For example, according to the Marshall Project, almost every state uses some form of "risk assessment" at some stage in the criminal justice system.

Risk assessments have existed in various forms for a century, but over the past two decades, they have spread through the American justice system, driven by advances in social science. The tools try to predict recidivism — repeat offending or breaking the rules of probation or parole — using statistical probabilities based on factors such as age, employment history, and prior criminal record. They are now used at some stage of the criminal justice process in nearly every state. Many court systems use the tools to guide decisions about which prisoners to release on parole, for example, and risk assessments are becoming increasingly popular as a way to help set bail for inmates awaiting trial.

This automated process relies on an algorithm in lieu of a judge's discretion. As noted by Cathy O'Neil, author of *Weapons of Math Destruction*, the data used by these algorithms to build models are sometimes suspect. Worse, we treat the output as "objective" without understanding that the data are themselves not objective. In this particular case, we set out to predict recidivism as if that means *the chance of committing a crime again* when in fact we are predicting *the chance of being arrested and convicted again*.

It does not take much imagination to see how being from a heavily policed area raises the chances of being arrested again, being convicted again, and in aggregate leads to even more policing of the same areas, creating a feedback loop. One can imagine similar issues with determining fit for a job, or credit-worthiness, or even face recognition and automated driving. In computing, we call this garbage-in-garbage-out: an algorithm is only as good as its data. This saying is certainly true, and especially relevant for AI algorithms that learn based on the data they are given.

Luckily, one way to address these issues is straightforward: to increase transparency. The kind of data the algorithm uses to build its model should be available. The decisions that such algorithms make should be inspectable. In other words, as we deploy these algorithms, each algorithm should be able to explain its output. “This applicant was assigned high risk because...” is more useful than, “This applicant was assigned high risk.”

If algorithms are inspectable, their creators are then able to call in outside experts to inspect them. After all, those with the knowledge to design an artificial intelligence algorithm can’t be expected to also be experts in medicine, the law, criminal justice, or banking. And outside experts shouldn’t have to get a Ph.D. in computer science to understand what programmers are doing with their data and their theories. AI transparency allows for a much wider range of input into any given project. And when things go wrong, it shows exactly where and how.

The idea of AI transparency is straightforward, but its implementation will be more complicated. First, the complexity of the algorithms makes it impractical for humans to inspect them manually. We will need tools that translate the complexity of AI algorithms into useable human-scaled insights.

Second, researchers have demonstrated that the more transparent an AI is, the easier it is to hack. Or worse still, if the AI is a trade secret, the easier it is to replicate. Therefore, we will also need new tools to secure every part of the programming and training process from unwanted intruders.

This does not mean that transparent AI is impossible, just that it presents a series of important technical challenges. But we must also recognize that transparency isn’t the only measure we can and should be taking to make AI responsible.

We also have the responsibility to consider the data sets that are used to train these algorithms. As shown in the earlier example about risk assessment for parolees, sometimes the data is skewed by the method that was used to collect it. This is a common problem in algorithms trained on social media data, to give another example.

Sometimes, the data set simply doesn’t contain enough information about underrepresented groups to even recognize them as a group. If that is the case, the data set can be expanded to include more information about those groups. Alternatively, they can add another “learner”

program to the AI that focuses on identifying those groups. This in and of itself presents a considerable challenge, however, because it suggests that the only way to make systems more responsible is to make them more complicated. To solve that problem, we need new concepts in computing theory to help us organize responsible AIs more efficiently. There is precedent for putting practice before theory; people wrote in code for thousands of years before the theory underlying modern public-key cryptography was laid out in the 1970s.

These technical problems present some of the major research challenges in artificial intelligence today. The National Institute of Standards and Technology's ongoing effort to create an AI risk management framework will need to incorporate these technical questions and others.

There are, of course, human issues as well. Right now, about 66 percent of tech workers are white, and 20 percent are Asian. Roughly 75 percent are men. Now, I work in AI, and I am not alleging that my colleagues are racist or misogynist. I am pointing out, however, that people from a subset of the population often build products that affect everyone. And often, they don't realize they're missing valuable perspectives.

In the long term, one of the key solutions to AI bias will be bringing a wider group of people into computing education, and into machine learning more specifically. We need to improve both the number and the diversity of people entering the field, starting from K-12 and extending to post-graduate work. One major obstacle is a lack of instructors at every level. In my own state, Georgia, only 35 percent of high schools that have AP programs offer AP Computer Science.

Now, K-12 isn't the only place for intervention, and programming is not the only job in artificial intelligence. In my own college, our DataWorks program trains unemployed adults to clean and integrate data sets for use in artificial intelligence projects. There are opportunities to open AI careers to more communities at every point in the pipeline.

While technical solutions are important, as are diversity and equity, a larger culture change is also needed. Computing has long been an intellectual Wild West, where things changed so fast that the priority was always to find the next, better solution. Now, we have succeeded in finding solutions so good that they are entwined in nearly every area of our personal lives and communities.

We have not as a field caught up to the reality of that responsibility. Unlike engineers or lawyers or medical professionals, we have not built responsibility for our actions into the structure of our field. We do of course have scholars specializing in ethical concerns. At Tech, that includes everything from autonomous robots in warfare to the relationship between software design and misinformation on social media.

I am not simply talking about ethics, or bias, or privacy, however, but instead a larger sense that computer scientists are responsible for how their products can be used or even abused. Our philosophy must catch up to the reality of our influence.

In conclusion, I am excited by this hearing. Advances in AI are central to our economic and social future. The issues are being raised here can be addressed with thoughtful support for robust funding in basic research in artificial intelligence—including research in AI transparency and new concepts in computing theory; support for AI education throughout the pipeline; and in developing standards for the responsible use of intelligent systems. These are all areas in which the funding power of the National Science Foundation and the National Institute of Standards and Technology can make a big difference.

I thank you very much for your time and attention today. I look forward to working with you in your efforts to understand how we can best develop these technologies to create a future where we are partners with intelligent machines.

Thank you. This concludes my testimony.

Dr. Charles Lee Isbell, Jr. received his B.S. in CS from the GeorgiaTech and his Ph.D. in CS from MIT. After four years at AT&T Labs/Research, he returned to Georgia Tech to join the faculty of the College of Computing. Charles' research interests are varied, but he is at heart a machine learning and artificial intelligence researcher. His recent work centers on building autonomous agents who engage in life-long learning when in the presence of thousands of other intelligent agents, including humans. Being human-centric, he finds himself studying the effects of AI bias. He and his work have been featured in the popular media as well as in technical collections. Charles also pursues reform in computing education focusing on broadening participation and access. He is an elected fellow of AAAI, ACM, and the American Academy of Arts and Sciences. In 2019, he assumed the role of the John P. Imlay, Jr. Dean for the College.

Chairwoman STEVENS. OK, Georgia Tech, you convinced me. I'm signing up for his class.

Dr. ISBELL. Done.

Chairwoman STEVENS. All right. With that, we're going to hear from Mr. Crenshaw for 5 minutes. Thanks.

**TESTIMONY OF MR. JORDAN CRENSHAW, VICE PRESIDENT
OF THE CHAMBER TECHNOLOGY ENGAGEMENT CENTER,
U.S. CHAMBER OF COMMERCE**

Mr. CRENSHAW. Thank you, Chair Stevens, Ranking Members Feenstra and Lucas, and Members of the Research and Technology Subcommittee. Good morning, and thank you. My name is Jordan Crenshaw, and I'm the vice president of the U.S. Chamber of Commerce's Technology Engagement Center. It's my pleasure to talk to you today about how we—business, government, and citizens—can work together to build trustworthy artificial intelligence.

AI is changing the world as we know it. By 2030, AI will have a \$16 trillion impact on the global economy. But from a practical level, what does that mean? AI is helping forecasters and emergency management better track the intensification of hurricanes and chart out evacuation and emergency preparedness. It's allowing researchers to more easily pinpoint virus mutations and tailor vaccines for new variants. It's also bolstering our cyber defenses against an evolving digital threat landscape. And finally, AI has the potential to fill the gaps where we have worker shortages, like patient monitoring where we have nursing shortages, and help tackle supply chain issues where we have a lack of available truckers.

The United States is not operating in a vacuum. Its strategic competitors also realize the benefits of this crucial technology. For example, prior to the invasion of Ukraine, China and Russia agreed to cooperate on developing emerging technologies, specifically noting artificial intelligence. When it comes to AI, we are in a race we must win. AI is here now, and it's not going away. We cannot ignore it, and we cannot afford to sit on the sidelines and allow those who do not share our democratic values to set the standard for the world.

For the research and deployment of AI to be successful, Americans must have trust in the technology. And while AI has many benefits, as I previously mentioned, in the wrong hands like those of our adversaries, there could be harms. Americans are united in the belief that we must beat our competitors as well. In fact, according to polling by the U.S. Chamber of Commerce, 85 percent of Americans believe the United States should lead in AI, and nearly that same number believes that we are best positioned as a nation to develop those ethical standards for its use.

We agree. It's why the Chamber earlier this year established its Commission on AI Competitiveness, Inclusion, and Innovation, led by your former congressional colleagues, Representatives John Delaney and Mike Ferguson, and it's comprised of experts in business, academia, and civil society. The Commission has been tasked with developing policy recommendations in three core areas: trustworthiness, work force preparation, and international competitiveness. Our Commission held field hearings in Austin, Silicon Valley,

Cleveland, London, and here in D.C.. And we've heard from a variety of stakeholders and look forward to presenting you with our recommendations early next year.

In the meantime, while we wait for the Commission to finalize its report, we offer the following observations about what it will take to maintain trustworthy AI leadership. The Federal Government has a significant role to play in conducting fundamental research in trustworthy AI. The Chamber was pleased to see passage of the *CHIPS and Science Act* and hopes to see the necessary appropriations to carry out the science provisions. We encourage continued investment in STEM (science, technology, engineering, and mathematics) education. We need a trained, skilled, and diverse work force that can bring together multiple voices for coding and developing systems.

AI is only as good, though, as the data it uses. That is why it is key that both government and the private sector team up to ensure there is quality data for more accurate and trustworthy AI. Governments should prioritize improving access to its own data and models and ways that respect individual privacy. At the same time, while we talk about privacy, as Congress looks to address these types of issues, it's important that we look at issues to determine whether or not we inhibit the collection of sensitive data and other types of data that could inhibit deploying trustworthy AI systems.

Fourth, we need to increase widespread access to shared computing resources. However, many small startups and academic institutions lack sufficient computing resources to help develop solutions to artificial intelligence. That's why Congress took the critical step of establishing the Research—passing the *Resource Task Force Act of 2020*. Now the National Science Foundation and the White House's Office of Science and Technology Policy should fully implement the law and expeditiously develop a roadmap to unlock AI innovation across multiple stakeholders.

Finally, we also are encouraged and are thankful for the work by NIST in its development of the AI Risk Management Framework, which is a consensus-driven, cross-sector, and voluntary framework to leverage best practices.

These recommendations are only the beginning. And I thank you for your time to address how the business community can partner with you to maintain trustworthy AI leadership. We thank you for your leadership, and I look forward to your questions.

[The prepared statement of Mr. Crenshaw follows:]



U.S. Chamber of Commerce

1615 H Street, NW
Washington, DC 20062-2000
uschamber.com

BEFORE THE U.S. HOUSE RESEARCH AND TECHNOLOGY SUBCOMMITTEE

Hearing on "Trustworthy AI: Managing the Risks of Artificial Intelligence"

Testimony of Jordan Crenshaw, Vice President, C_TEC, U.S. Chamber of Commerce

September 29, 2022

Dear Chairman Stevens, Ranking Member Feenstra, and distinguished Research and Technology Committee members. First, thank you for your invitation to come before you today to testify. My name is Jordan Crenshaw, and I am honored to serve as the Vice President of the U.S. Chamber Technology Engagement Center (C_TEC) at the U.S. Chamber of Commerce. C_TEC is the technology hub within the U.S. Chamber, and our goal is to promote the role of technology in our economy and advocate for rational policy solutions that drive economic growth, spur innovation, and create jobs. Today's hearing titled "Trustworthy AI: Managing the Risks of Artificial Intelligence" is a timely and critical discussion, and the Chamber appreciates the opportunity to participate.

The world has quickly entered its fourth industrial revolution, in which the use of technology and artificial intelligence ('AI') is helping propel humanity. However, we are witnessing the benefits of using AI daily, from its value in adapting vaccines to tailor them to new variants to increasing patient safety during procedures like labor and delivery.¹ Artificial intelligence is also rapidly changing how businesses operate. This emerging technology is a tremendous force for good in its ability to secure our networks, expand opportunities for the underserved, and make our communities safer and more prosperous.²

America is currently in a race with countries like China to lead in Artificial Intelligence.³ America's competitors may not respect the same values as our allies, such as individual liberties, privacy, and the rule of law. While the development and deployment of AI have become an essential part of facilitating innovation, this innovation will never reach its full potential and enable the United States to compete without trust. The business community understands that fostering this trust in AI technologies is essential to advance its responsible development, deployment, and use. This has been a core understanding of the U.S. Chamber, as it is the first principle within the 2019 "U.S. Chamber's Artificial Intelligence Principles:

¹ <https://www.5newsonline.com/article/news/health/northwest-health-introducing-new-technology-to-enhance-maternal-and-fetal-safety/527-9c173d18-c56e-457b-8317-62ebaae93558>

² <https://americaninnovators.com/research/data-for-good-promoting-safety-health-and-inclusion/>

³ <https://www.washingtonpost.com/opinions/2022/09/13/artificial-intelligence-ai-high-tech-race-with-china/>

Trustworthy AI encompasses values such as transparency, explainability, fairness, and accountability. The speed and complexity of technological change, however, mean that governments alone cannot promote trustworthy AI. The Chamber believes that governments must partner with the private sector, academia, and civil society when addressing issues of public concern associated with AI. We recognize and commend existing partnerships that have formed in the AI community to address these challenges, including protecting against harmful biases, ensuring democratic values, and respecting human rights. Finally, any governance frameworks should be flexible and driven by a transparent, voluntary, and multi-stakeholder process.⁴

AI also brings a unique set of challenges that should be addressed so that concerns over its risks do not dampen innovation and to help ensure the United States can lead globally in trustworthy AI. The U.S. Chamber of Commerce's Technology Engagement Center (C_TEC) shares the perspective with many of the leading government and industry voices, including the National Security Commission on Artificial Intelligence (NSCAI)⁵, the National Institute of Standards and Technology (NIST)⁶, that government policy to advance the ethical development of AI-based systems, sometimes called "responsible" or "trustworthy" AI, can enable future innovation and help the United States to be the global leader in AI.

This is why we have prioritized the need to build public trust in AI through our continued efforts. The U.S. Chamber earlier this year launched its Artificial Intelligence (AI) Commission on Competition, Inclusion, and Innovation to advance U.S. leadership in using and regulating AI technology.⁷ The Commission, led by co-chairs former Congressmen John Delaney and Mike Ferguson, is composed of representatives from industry, academia, and civil society to provide independent, bipartisan recommendations to aid policymakers with guidance on artificial intelligence policies as it relates to regulation, international research, development competitiveness, and future jobs.

Over the past few months, the Commission has heard oral testimony from 87 expert witnesses⁸ over five separate field hearings. The Commission heard from individuals such as Jacob Snow, Staff Attorney for the Technology & Civil Liberties Program at the ACLU of Northern California. In his testimony, he told the Commission that the critical discussions on AI are "not narrow technical questions about how to design a product. They are social questions about what happens when a product is deployed to a society, and the consequences of that deployment on people's lives."⁹

Doug Bloch, Political Director at Teamsters Joint Council 7, referenced his time serving on Governor Newsom's Future of Work Commission: "I became convinced that all the talk of

⁴ <https://www.uschamber.com/technology/us-chamber-releases-artificial-intelligence-principles>

⁵ <https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf>

⁶ <https://www.nist.gov/artificial-intelligence>

⁷ www.americaninnovators.com/aicommission

⁸ <https://americaninnovators.com/aicommission/>

⁹ <https://americaninnovators.com/news/ai-for-all-experts-weigh-in-on-expanding-ais-shared-prosperity-and-reducing-potential-harms/>

the robot apocalypse and robots coming to take workers' jobs was a lot of hyperbole. I think the bigger threat to the workers I represent is the robots will come and supervise through algorithms and artificial intelligence."¹⁰

Miriam Vogel, President and CEO of EqualAI and Chair of NAIAC, also addressed the Commission. She stated "I would argue that it's not that we need to be a leader, it's that we need to maintain our leadership because our brand is trust."

The Commission also received written feedback from stakeholders answering numerous questions that the Commission has posed in three separate requests for information (RFI), which asked questions about issues ranging from defining AI, balancing fairness and innovation,¹¹ and AI's impact on the workforce.¹² These requests for information outline many of the fundamental questions that we look to address in the Commission's final recommendations, which will help government officials, agencies, and the business community. The Commission is diligently working on its recommendations and will look to release them earlier next year.

While the Chamber is diligently taking a leading role within the business community to address many of the concerns which continue to be barriers to public trust and consumer confidence in the technology, my testimony before you today will look to address the following underlying questions:

- What are the opportunities for the federal government and industry to work together to develop trustworthy AI?
- How are different industry sectors currently mitigating risks that arise from AI?
- How can the United States encourage more organizations to think critically about risks that arise from AI systems, including ways in which we prioritize trustworthy AI from the earliest stages of development of new systems?
- How can the federal government strengthen its role in the development and responsible deployment of trustworthy AI systems?

I. Opportunities for the Federal Government and Industry to Work Together to Develop Trustworthy AI

A. Congress Needs to Pass a Preemptive National Data Privacy Law

Artificial Intelligence relies upon the data in which it is provided. Particularly sensitive data can be used to determine whether AI systems operate fairly. Many underlying concerns regarding the use of Artificial Intelligence will need to be reassessed should a National Data

¹⁰ <https://americaninnovators.com/news/ai-for-all-experts-weigh-in-on-expanding-ais-shared-prosperity-and-reducing-potential-harms/>

¹¹ https://americaninnovators.com/wp-content/uploads/2022/04/CTEC_RFI-AIcommission_2.pdf?utm_source=sfmc&utm_medium=email&utm_campaign=&utm_term=RFI+3+-+Workforce+-+20220518&utm_content=5/19/2022

¹² https://uschambermx.iad1.qualtrics.com/jfe/form/SV_cMw5ieLrIsFwUPs

Privacy bill be signed into law as new well-defined rules are put into place. The U.S. Chamber has been at the forefront of advocating for a true *national* privacy standard that gives strong data protections for all Americans equally. For this reason, the Chamber was the first trade association after the passage of the California Consumer Privacy Act to formalize and propose privacy principles and model legislation.¹³

Most central to a national privacy law is the need for true preemption that creates a national standard. A patchwork of fifty different state laws¹⁴ would eliminate the certainty required for data subjects and businesses in compliance and operations. According to a recent report from ITI, a fifty-state patchwork of comprehensive privacy laws could cost the economy \$1 trillion and \$200 million for small businesses.¹⁵ Recently, the Chamber released findings that nearly 25 percent of small businesses plan to use artificial intelligence and 80 percent of these businesses believe limiting access to data would harm their operations.¹⁶ A state patchwork exacerbates the difficulties these businesses face.

Recently, the House Energy and Commerce Committee reported the American Data Privacy and Protection Act (“ADPPA”).¹⁷ Although the ADPPA has many laudable consumer protections like the right to delete, opt out of targeted advertising, as well as data correction and access, there are significant concerns that it could create a new national patchwork and cut off access to data which could improve AI fairness.¹⁸ For example, the bill would only preempt what is covered by the Act and would empower the FTC to bar the collection and use of data.

We encourage stakeholders and Congress to work together to pass a truly preemptive privacy law that enables the use of data to improve AI—not inhibit the deployment of AI.

B. Support for Alternative Regulatory Pathways Such As Voluntary Consensus Standards

New regulation is not always the answer for emerging or disruptive technologies. Non-regulatory approaches can often serve as effective tools to increase safety and build trust, and allow for flexibility and innovation. This is particularly applicable to emerging technologies such as artificial intelligence as the technology continues to rapidly evolve.

This is why the Chamber supports the National Institutes of Science and Technology’s (NIST) work in drafting the Artificial Intelligence Risk Management Framework (AI_RMF). The

¹³ <https://www.uschamber.com/technology/data-privacy/the-10-principles-of-data-privacy>

¹⁴ https://americaninnovators.com/wp-content/uploads/2022/01/CTEC_Privacy2022_HeatMap-1024x791-1.pdf

¹⁵ <https://itif.org/publications/2022/01/24/50-state-patchwork-privacy-laws-could-cost-1-trillion-more-single-federal/>

¹⁶ <https://americaninnovators.com/wp-content/uploads/2022/08/Empowering-Small-Business-The-Impact-of-Technology-on-U.S.-Small-Business.pdf>

¹⁷ <https://docs.house.gov/meetings/IF/IF00/20220720/115041/BILLS-117-8152-P000034-Amdt-1.pdf>

¹⁸ <https://www.uschamber.com/technology/data-privacy/what-should-and-should-not-be-included-in-a-national-privacy-bill>

AI RMF is meant to be a stakeholder-driven framework, which is “intended for voluntary use and to improve the ability to incorporate trustworthiness considerations into the design, development, use, and evaluation of AI products, services, and systems.”

Another example of non-regulation tools is the National Highway Traffic Safety Administration’s (“NHTSA”) Voluntary Safety Self-Assessments (“VSSA”). More than two dozen AV developers have submitted a VSSA to NHTSA, which have provided essential and valuable information to the public and NHTSA on how developers are addressing safety concerns arising from AVs. The flexibility provided by VSSAs, complemented by existing regulatory mechanisms, provides significant transparency into the activities of developers without compromising safety.

Voluntary tools provide significant opportunities for consumers, businesses, and the government to work together to address many of the underlying concerns with emerging technology while at the same time providing the necessary flexibility to allow the standards not to stifle innovation. These standards are pivotal in the United States’ ability to maintain leadership in emerging technology as it is critical to ensuring our global economic competitiveness in this cutting-edge technology.

C. Stakeholder Driven Engagement

The U.S. Chamber of Commerce stands by and is ready to assist the government in any opportunity to improve consumer confidence and trust in AI systems. We have always viewed trust as a partnership, and only when government and industry work side by side can that trust be built. The opportunities to facilitate this work are great, but there are essential steps that industry and government can make today.

We asked the American public earlier this year about their perception of artificial intelligence. The polling results were very eye-opening, as there was a significant correlation between the trust and acceptance of AI and an individual’s knowledge and understanding of the technology.¹⁹ To build the necessary consumer confidence to allow artificial intelligence to grow for the betterment of all, all opportunities must be taken advantage of for industry and governments to work together in educating stakeholders about the technology.

This is why we appreciate the National Institute of Science and Technology’s (NIST) work in drafting the Artificial Intelligence Risk Management Framework (AI_RMF). The AI RMF is meant to be a stakeholder-driven framework. NIST’s continued engagement with all stakeholders in the development of the framework is important to develop trust between government and industry. To date, NIST actions include two workshops, with a third workshop scheduled for next month. They have also included three engagement opportunities for stakeholders to provide written feedback on the development, direction, and critique of the AI

¹⁹ <https://americaninnovators.com/wp-content/uploads/2022/01/CTEC-US-Outlook-on-AI-Detailed-Analysis.pdf>

RMF. This engagement by NIST has allowed for the development of trust between industry and the federal government. While we implore NIST and their action on the RMF, it's prudent to highlight that NIST is only one entity within the federal government and that other agencies and regulators should look to the model.

D. Awareness of the Benefits of Artificial Intelligence

At the same time, it is critical that federal agencies do not seek to prescriptively regulate technologies without first establishing a strong public record. The business community has significant concerns about the Federal Trade Commission undertaking rulemaking on privacy, security, and algorithms asking whether it should make economy-wide rules on algorithmic decision systems.²⁰ First and foremost, the FTC should enable NIST's process to conclude and let Congress speak clearly about how it wants to make policy in artificial intelligence before undertaking general rulemaking. "NIST contributes to the research, standards, and data required to realize the full promise of artificial intelligence (AI) as a tool that will enable American innovation, enhance economic security and improve our quality of life."²¹ Therefore, we believe it is essential for NIST to be able to finish the RMF to provide the necessary robust record within the federal government. This is vital, as government agencies such as the FTC ask technical questions.

E. Awareness of the Benefits of Artificial Intelligence

Another excellent opportunity for industry and government to work together is highlighting the benefits and efficiencies of the use of technology within the government. The government's utilization of AI has the ability to lead to medical breakthroughs²² to help to predict risk for housing and food insecurities.²³ AI is helping our government provide better assistance to the American public, and is becoming a vital tool. The development of these resources does not come in a vacuum, and the majority of these tools are done so in partnership with industry. Highlighting these workstreams and the benefits that they deliver for the American public can assist in fostering trust in technology, as well as build overall consumer confidence in technology use outside of government.

However, this would also require a foundational change in how our government works, which includes addressing the "legacy culture" that has stifled the necessary investment and buildout of 21st-century technology solutions and harnessing data analytics. Congress's passage of the Modernizing Government Technology Act during the 115th Congress was an essential first step in rectifying decades of needed investment. However, this legislation, while important, will not alone fix the problem and would ask Congress to continue to do necessary

²⁰ <https://thehill.com/opinion/technology/3621149-the-ftc-needs-a-reminder-that-its-a-regulator-not-a-legislator/>

²¹ <https://www.nist.gov/artificial-intelligence>

²² <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/deloitte-analytics/us-ai-institute-government-public-services-dossier.pdf>

²³ <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/deloitte-analytics/us-ai-institute-government-public-services-dossier.pdf>

oversight within the federal government IT sector so that the essential and sustained investments can be made.

II. How are Different Sectors Adopting Governance Models and Other Strategies to Mitigate Risks that Arise from AI Systems?

AI is a tool and does not exist in a legal vacuum. Policymakers should be mindful that activities performed and decisions aided by AI are often already accountable under existing laws. Where new public policy considerations arise, governments should consider maintaining a sector-specific approach while removing or modifying those regulations that act as a barrier to AI's development, deployment, and use. In addition, governments should avoid creating a patchwork of AI policies at the subnational level and should coordinate across governments to advance sound and interoperable practices.

It's also important to highlight that there is a market incentive for companies to address associated risks with the use of artificial intelligence. Companies to begin with are very risk-averse when it comes to potential legal liabilities associated with their use of the technology. This is why we applaud NIST's development of "Playbook," which is "designed to inform AI actors and make the AI RMF more usable."²⁴ We believe the playbook will provide a great resource for the business community and industry in helping them evaluate risk.

Every sector will have different risks associated with the use of AI, which is why it is important to maintain a sector-specific approach. However, we believe it's important for policy makers to do necessary oversight to close current legal gaps. For this reason, we would ask policy makers to do necessary oversight of the American COMPETE Act, which requires the U.S. Department of Commerce and Federal Trade Commission ('FTC') to look at different emerging technologies and to conduct a thorough analysis of current standards, guidelines, and policies regarding AI that are implemented by each government agency, as well as industry-based bodies. This important assessment would provide lawmakers and industry with a comprehensive and baseline understanding of relevant regulations that are already in place.

III. How Should the United States Encourage More Organizations to Think Critically about Risks that Arise from AI Systems, Including by Prioritizing Trustworthy AI from the Earliest Stages of Development of New Systems?

The United States has a great opportunity through the development of the NIST AI RMF to provide organizations with a key set of documents that would assist in their ability to think critically about risk. The adaptable voluntary framework would assist companies from big to small in assessing the risk with which they are comfortable and provide guidance on ways to help critical thinking through potential negative externalities which may be associated with its use. That being said, the framework can only assist if it is in the hands of those creating and developing and those who oversee its use. For this reason, we believe that NIST and the

²⁴ <https://www.nist.gov/itl/ai-risk-management-framework/nist-ai-rmf-playbook-faqs>

Department of Commerce should look at ways in which they can reach all different demographics and stakeholders to make them aware of these resources.

Furthermore, we believe further effort should be made by the government to make connections to those small and medium size businesses that usually lack the time and resources to be looking for things like the RMF.

IV. **What Recommendations do you Have for how the Federal Government can Strengthen its Role for the Development and Responsible Deployment of Trustworthy AI Systems?**

The federal government has the ability to take a leading role in strengthening the development and deployment of artificial intelligence. We believe that the following recommendations should be acted on now.

First, we would advise the federal government to conduct fundamental research in trustworthy AI: The federal government has played a significant role in building the foundation of emerging technologies through conducting fundamental research. AI is no different. A recent report that the U.S. Chamber Technology Center and the Deloitte AI Institute²⁵ surveyed business leaders across the United States had 70% of respondents indicated support for government investment in fundamental AI research. The Chamber believes that the CHIPS and Science Act was a positive step in the necessary investment, as the legislation authorizes \$9 Billion for the National Institutes of Standards Technology (NIST) for Research and Development and advancing standards for "industries of the future," which includes artificial intelligence. Furthermore, we have been a strong advocate for the National Artificial Intelligence Initiative Act, which was led by Chairwoman Eddie Bernice Johnson and Ranking Member Lucas, which developed the office of the National AI Initiative Office (NAIO) to coordinate the Federal government's activities, including AI research, development, demonstration, and education and workforce development.²⁶ We would strongly advise members to appropriate these efforts fully.

Second, we encourage continued investment into Science, Technology, Engineering, and Math Education (STEM). The U.S. Chamber earlier this year polled the American public on their perception of artificial intelligence. The findings were clear; the more the public understands the technology, the more comfortable they become with its potential role in society. We see education as one of the keys to bolstering AI acceptance and enthusiasm as a lack of understanding of AI is the leading indicator for a push-back against AI adoption.²⁷

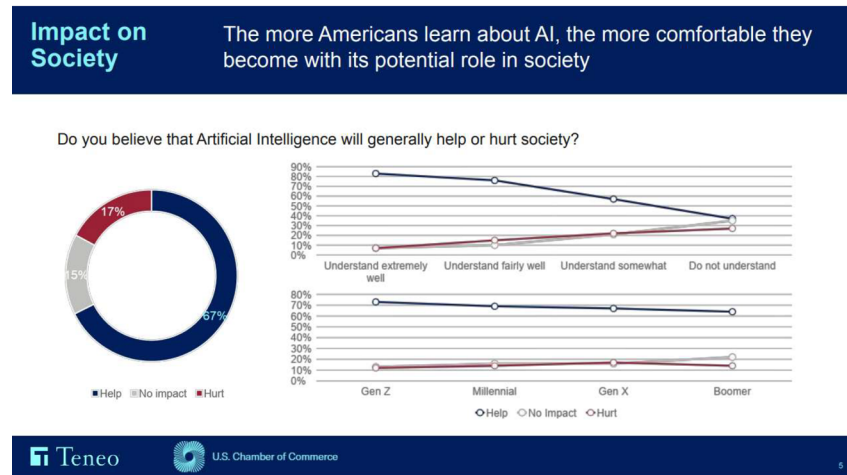
The Chamber strongly supported the CHIPS and Science Act, which made many of these critical investments, including \$200 million over five years to the National Science

²⁵ <https://www.uschamber.com/technology/investing-trustworthy-ai>

²⁶ <https://www.ai.gov/naio/>

²⁷ <https://americaninnovators.com/wp-content/uploads/2022/01/CTEC-US-Outlook-on-AI-Detailed-Analysis.pdf>

Foundation (NSF) for domestic workforce build-out to develop manufacture chips, and also \$13 Billion to the National Science Foundation for AI Scholarship-for-service. However, the authorization within the legislation is just the start; we now ask Congress to appropriate the funding for these important investments.



Third, the government should prioritize improving access to government data and models: High-quality data is the lifeblood of developing new AI applications and tools, and poor data quality can heighten risks. Governments at all levels possess a significant amount of data that could be used to improve the training of AI systems and create novel applications. When C_TEC asked leading industry experts about the importance of government data, 61% of respondents agree that access to government data and models is important. For this reason, we would encourage policymakers to build upon the success of the OPEN Government Data Act by providing further additional funding and oversight to allow for expanding the scope of the law to include non-sensitive government models as well as datasets at the state and local levels.

Fourth, Increase widespread access to shared computing resources : In addition to high-quality data, the development of AI applications requires significant computing capacity. However, many small startups and academic institutions lack sufficient computing resources, which in turn prevents many stakeholders from fully accessing AI's potential. When we asked stakeholders within the business community about the importance of shared computing capacity, 42% of respondents supported encouraging shared computing resources to develop and train new AI models. Congress took a critical first step by enacting the National AI Research Resource Task Force Act of 2020. Now, the National Science Foundation and the

White House's Office of Science and Technology Policy should fully implement the law and expeditiously develop a roadmap to unlock AI innovation across all stakeholders.

Fifth, Enable open source tools and frameworks : Ensuring the development of trustworthy AI will require significant collaboration between government, industry, academia, and other relevant stakeholders. One key method to facilitate collaboration is through encouraging the use of open source tools and frameworks to share best practices and approaches to trustworthy AI. An example of how this works in practice is the National Institute of Standards and Technology's (NIST) AI Risk Management Framework (RMF), which is intended to be a consensus-driven, cross-sector, and voluntary framework, akin to NIST's existing Cybersecurity Framework, whereby stakeholders can leverage as a best practice to mitigate risks posed by AI applications. Policymakers should recognize the importance of these types of approaches and continue to support their development and implementation

V. Conclusion

AI leadership is essential to global economic leadership in the 21st century. According to one study, AI will have a \$13 trillion impact on the global economy by 2030.²⁸ Through the right policies, the federal government can play a critical role in incentivizing the adoption of trustworthy AI applications. The United States has an enormous opportunity to transform the economy and society in positive ways through leading in AI innovation as other economies contemplate their approach to trustworthy AI forward on how U.S. policymakers can pursue a wide range of options to advance trustworthy AI domestically and empower the United States to maintain global competitiveness in this critical technology sector. The United States must be the global leader for AI trustworthiness for the technology to develop in a manner that is balanced and takes into account basic values and ethics. The United States can only be a global leader if the administration and Congress work together on a bipartisan basis. We are in a race we can't afford to lose.

²⁸ <https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy>

Jordan Crenshaw serves as Vice President and leads the day-to-day operations at the U.S. Chamber of Commerce's [Technology Engagement Center](#). Crenshaw also directly manages the Chamber's privacy working group which is comprised of nearly 300 companies and trade associations, which developed model privacy legislation and principles. Prior to becoming vice president of C TEC, he led the Chamber's Telecommunications and E-Commerce Policy Committee, which analyzes federal privacy, cloud computing, broadband, internet, e-commerce, and broadcast policies that impact U.S. businesses.

Before joining the Chamber, Crenshaw served as an attorney focusing on environmental issues and analysis of consumer privacy laws. Crenshaw also worked at McGuireWoods, LLP assisting discovery issues for environmental nuisance, TCPA, and other civil litigation. Crenshaw also served Virginia Senate leadership, the Office of the Attorney General of Virginia, the U.S. Department of Labor Office of Administrative Law Judges, and the National Right to Work Defense Foundation.

Crenshaw earned both his undergraduate degree and Juris Doctor from the College of William and Mary. He is licensed to practice law in Virginia and is a Certified Information Privacy Professional (CIPP/US). He and his wife Molly live in Virginia.

Chairwoman STEVENS. Thank you.
With that, Ms. Singh, yes.

**TESTIMONY OF MS. NAVRINA SINGH,
FOUNDER AND CHIEF EXECUTIVE OFFICER, CREDO AI**

Ms. SINGH. Madam Chair, Ranking Member Feenstra and Lucas, and Members of the Subcommittee, thank you for the opportunity to testify today and to be part of this distinguished panel of witnesses. My name is Navrina Singh. I'm the Founder and CEO of Credo AI, a venture-backed startup. In addition, I'm a member of the National AI Advisory Committee that is advising President Biden as part of the National AI Initiative.

Trustworthy artificial intelligence is a topic that is deeply personal to me. Growing up in India as a girl who aspired to be an engineer, I learned early on that I faced an uphill battle for no reason other than my gender. Part of my passion for the subject and the main reason I founded Credo AI in March 2020 is because I experienced firsthand what is at stake. While AI is an exciting and ultimately very useful technology, unless we create a culture of accountability, transparency, and governance around it, we risk unchecked growth and algorithms that may unintentionally encode the same types of societal ceilings and perceptions that I experienced as a girl in India and that many others still experience today.

Members of the Subcommittee know very well the power and potential of AI when used responsibly. While it is a transformational technology that is evolving rapidly, I realize that there are different points of view on its perceived advantages. But one thing we can all agree on is AI is not going away, which is why we owe it to ourselves and to the world that our children will inherit to ensure robust compliance and governance structures to keep pace with the AI development.

As the Subcommittee studies the question of how to manage AI risk and build trustworthy AI, we think three key considerations merit special attention. First, I want to focus on full AI lifecycle, from design to development, to testing and validation, to production and use. That means building AI systems responsibly continuously. It is fit for purpose, fair, transparent, safe and secure, privacy-preserving, and auditable.

Second, context is paramount. We believe that achieving trustworthy AI depends on shared understanding, that governance and oversight of AI is industry-specific, application-specific, model-specific, and data-specific to ensure that it is fit for purpose. This necessitates a collaborative approach to metric alignment, and associated assessments.

Third, transparency reporting and system assessments are critical for responsible AI governance. Reporting requirements that promote and incentivize public disclosure of AI system behaviors act as a key driver for establishment of standards and benchmark. And fundamental to this is access to compliant and comprehensive data for assessments. For these reasons, we at Credo AI advocate for context base, full AI lifecycle governance of AI systems with reporting requirements that are specific, regular, and transparent.

If you truly want to be a global leader in AI, then our focus should be on building responsible technology aligned with our societal values. Responsible AI is also a competitive advantage. It allows companies to deploy AI at scale with confidence, and this transparency promotes trust with consumers in this technology. Government has a critical role to play here, working together through public-private partnerships to ensure the right set of standards exist to further innovation in the space. And we urge the policymakers and standard-setting bodies to prioritize establishing context-focused standards and benchmarks that are globally interoperable and can help eliminate some of the guesswork.

My 8-year-old daughter told me recently that she wants to be an inventor and a social media influencer when she grows up. While I'm grateful that in this country my daughter will have the opportunity to follow her dreams, we owe it to her and the generations that will follow to ensure that we build AI which is developed responsibly and ethically.

Thank you for the opportunity to appear before you, and I look forward to your questions.

[The prepared statement of Ms. Singh follows:]

**PREPARED TESTIMONY OF NAVRINA SINGH, FOUNDER AND CEO, CREDO AI
BEFORE THE HOUSE COMMITTEE ON SCIENCE, SPACE AND TECHNOLOGY
SUBCOMMITTEE ON RESEARCH AND TECHNOLOGY**

**HEARING DATE/TIME: SEPTEMBER 29, 2022 10:30 A.M. EST
HEARING TITLE: TRUSTWORTHY AI: MANAGING THE RISKS OF ARTIFICIAL
INTELLIGENCE**

Introduction

Madam Chair, Ranking Member Feenstra, and Members of the Subcommittee on Research and Technology, thank you for the opportunity to testify today and to be a part of this distinguished panel of witnesses. My name is Navrina Singh, and I am the Founder and Chief Executive Officer of Credo AI.

A “credo” is a statement or system of beliefs or principles to guide actions. I founded Credo AI in March 2020 with one key goal in mind: to enable organizations to deliver Responsible AI (RAI) at scale. RAI includes assessment, reporting, and governance to the highest of ethical standards in order to ensure a fair, transparent, compliant, and auditable environment for the development and use of artificial intelligence (AI).

Credo AI is a software company, and our core product is the Credo AI Responsible AI Governance Platform™. Our platform is designed to help organizations consistently translate principles into actionable metrics, assessments and benchmarks throughout the entire AI lifecycle. We recognize that enterprises are at different levels of maturity in their development and use of AI. Our mission at Credo AI is to realize comprehensive Responsible AI governance by providing software tools to enterprises wherever they are in their AI Governance journey.

What Is Responsible AI?

At Credo AI, we define the phrase “Responsible AI” as AI that is human centered. That means that AI systems need to be performant, fair, transparent, safe and secure, privacy-preserving, and auditable. These tenets are aligned with the ways that many other organizations, regulatory bodies, and standard-setting bodies define the phrase “Responsible AI.” Our customers aren’t only concerned with making sure their systems are accurate or performant; they want to know if their systems are fair, transparent, and robust, and they want to know how regulators are defining these parameters. Measuring and managing each of these tenets is very complex and context-dependent—each must be aligned, based on the context of their use, with the values both of society and the organization developing or using the AI.

Credo AI’s RAI Governance Platform is built to help organizations map, measure, manage, and mitigate AI risk and compliance for all their AI use cases. The Platform provides organizations with context-driven requirements for their AI use cases in the form of “Policy Packs.” Policy Packs provide specific technical and process requirements that an AI system must meet, based on regulations, laws, standards, frameworks, guidelines, an organization’s internal guardrails, and industry best practices. Our platform connects to our open source Responsible AI assessment framework, Credo AI Lens, which can be used to assess machine learning (ML) models and datasets based on the requirements coming from Policy Packs, allowing our customers to programmatically generate governance artifacts like model cards, assessments reports,

transparency reports or audit reports. The platform standardizes governance activities, promotes multidisciplinary collaboration among technical and business stakeholders, and reduces the burden of governance on technical teams, making it easier for organizations to govern their AI systems more effectively and gain confidence in their AI use.

How to Create an Environment that Fosters RAI

AI is a transformative technology that is rapidly evolving. There is a significant opportunity to encourage the development of trustworthy technology and set effective policy, and as the Subcommittee studies this important issue, Credo AI respectfully offers the following key points for your consideration:

- **RAI Requires a Full Lifecycle Approach:** At Credo AI, we believe managing risk is only one part of delivering on the promise of Responsible AI—it demands a full lifecycle approach. AI systems cannot be considered “responsible” based on one point-in-time snapshot, but instead must be continuously evaluated for responsibility, and transparently reported on throughout the entire AI lifecycle, from the design, to development, to testing and validation, to production and use.
- **There Is No One-Size-Fits-All Approach to AI Governance:** We believe that achieving trustworthy AI depends on a shared understanding that AI is industry specific, application specific, data specific and context driven. There is no one-size-fits-all approach to “what good looks like” for most AI use cases. For example: there is no single definition of algorithmic “fairness,” because the concept of fairness is incredibly context-dependent. Similarly, when considering what metric or measures to use for the performance of an AI system, assessors should be able to select from a wide variety of different metrics that take into account use case context, model type, and data type. The organization building the AI system should be consulted about “acceptable” performance metrics. This requires a collaborative approach to assessments, and we advocate for context-based tests for AI systems with reporting requirements that are: specific, regular, and transparent.
- **Transparency Reporting and System Assessments Can Deliver Trustworthy and Accurate AI:** The importance of transparency reporting and system assessments cannot be overstated as a critical foundation for RAI governance for all organizations. Reporting allows policymakers to start to evaluate different approaches, and potentially opens the door for benchmarking—reporting is the step that gets us to standards that can be enforced. We have seen firsthand how comprehensive and accurate assessments of the AI applications and the associated models/datasets, coupled with transparency and disclosure reporting, encourage responsible practices to be cultivated, engineered, and managed throughout the AI development life cycle. Fundamental to this is access to compliant and comprehensive data for assessments.

Companies Are Seeking Guidance

In our experience, organizations understand that Responsible AI is a competitive advantage for them in this age of AI. Organizations know there is a need for RAI governance, and welcome a collaborative approach to developing it. The notion that AI regulation will cause U.S. companies to offshore or cause AI to stagnate is a *false* premise. Based on our experience in the field working with companies that develop and deploy AI, we repeatedly hear a desire to have those systems work well in a compliant, safe, fair and auditable fashion. This leads to an important synergy: the more that policymakers can do to help companies understand how to develop trustworthy systems, the easier it will be for those companies to maximize the value of those systems. Thoughtful policy making and governance via public-private partnership can create conditions for innovations in AI for these companies.

Key Challenges to Overcome in the Development and Use of Responsible AI

While there is reason for optimism, there is much work to be done. Credo AI has experience working with customers across industries, and we have observed that they are all working to set up processes to foster RAI. The key challenges that we have observed and that we hope policymakers will consider when it comes to more effectively promoting the responsible development and use of AI include:

- **Standards and benchmarks for RAI are still emerging.** We urge policymakers and standard-setting bodies to prioritize establishing context-focused standards and benchmarks—that are globally interoperable—that can help take some of the guesswork out of compliance with AI regulations. While many emerging regulations set “fairness,” “transparency,” and other RAI dimensions as key requirements for compliance, there are not yet clear standards or benchmarks for what it means for an AI system to be “fair” or “transparent.” That is because there are many ways to define these terms. Without clear standards and benchmarks, organizations are left having to develop and justify their own measures for different technical dimensions of their AI systems. Standards and benchmarks should also try to account for the challenges of operationalizing such requirements and frameworks depending on the size and reach of the organization. Expecting a small or mid-sized business to operationalize new standards as quickly as major multinational companies would present its own challenges.
- **AI regulations must include reporting requirements to foster transparency and drive towards standards.** We urge policy makers to establish requirements that mandate disclosures and transparency reporting around the procurement, development, and use of AI. Because of the lack of standards today, many organizations are reluctant to share results about the behavior of their AI systems externally—because they have no idea how their results might compare with those of their competitors, or whether they are “good” or “bad” for external stakeholders. We are strong supporters of reporting requirements,

therefore, that promote and incentivize public disclosure of AI system behavior and operation as a key driver of the establishment of standards and benchmarks.

Context is Critical: Metrics for Each Tenant of RAI Vary

We strongly believe that AI is industry-specific, application-specific, and context-driven and needs to continuously assessed —factors that should be reflected in its governance.

For example, when considering what definition to use for fairness, we feel that there is no one-size-fits-all answer or approach. There is not a single definition of algorithmic fairness accepted across industry sectors and use cases.

Algorithmic fairness is a field of research aimed at understanding and correcting the ways that historical societal biases show up in AI systems. An AI system can be considered to be “fair,” in the sense of algorithmic fairness, if it does not perpetuate or amplify harmful societal biases in its operation.

When data scientists are evaluating whether their AI systems are fair, they look at specific technical measures of *bias* in their AI systems—to understand if these systems are perpetuating harmful societal biases. There are two primary ways that we measure bias in our AI systems: evaluating **parity of performance** and **parity of outcomes**.

- Parity of performance is about evaluating whether your ML model performs equally *well* for all different groups that interact with it. For example, does your facial recognition system detect Black women’s faces at the same or similar accuracy rate that it detects white men’s faces?
- Parity of outcomes is about evaluating whether your ML model confers a benefit to different groups at the same rate. For example, does your candidate ranking system recommend Black women get hired at the same or similar rate as it recommends white men?

We do not have a singular definition of fairness — nor should anyone who is thinking about algorithmic fairness — because fairness is incredibly context-dependent.

Here’s an example to illustrate why you cannot have a “one size fits all” definition of algorithmic fairness. Let’s say that you have an AI system that is going to be predicting whether someone should be given a loan (a credit risk prediction system), and you have another AI system that is going to predict whether somebody has cancer by analyzing a CT scan for tumors. For your credit risk prediction system, the system is considered “unfair” if it predicts that Black women are credit-worthy (and therefore should be given a loan) at a much lower rate than white men; we want to make sure that our credit prediction system is conferring the benefit of getting a loan

relatively equally across groups, regardless of gender or race. This is an example of parity of outcomes. For the cancer detection system, however, the parity of outcomes isn't the primary concern; we don't care if the system is predicting that women have breast cancer at a rate that is significantly higher than men. This is because for this cancer detection system to be considered "fair," we want to make sure that it is *equally accurate* for all groups that interact with it. The issue here is parity of performance: our cancer detection system will be considered fair if it has the same performance rate across all groups.

The metrics that you use to measure parity of performance are different from the metrics that you use to measure parity of outcomes—and even within these two categories, there are many different metrics that you can pick, depending on what is most important based on the use case context.

Similarly, when considering the question of what metric or measure to use for algorithmic performance: there is no single metric for performance. Depending on your use case context, model type, and data type, you may select from a wide variety of different metrics that are all reasonable and accepted ways to evaluate performance of an AI/ML system.

For a cancer detection system, assessors might care more about a system that has relatively equal *false negative rates* across groups, because incorrectly diagnosing someone as healthy who actually has cancer is a life-threatening mistake (the cost of making an incorrect "negative" prediction is very high). For a facial recognition system that is going to be used to grant access to a device—say, your phone—assessors may care more about *false positive rate*, however, because they want to ensure that this system doesn't accidentally grant access to your phone to someone who should not have access (the cost of making an incorrect "positive" prediction is very high).

These examples are all intended to show that there is no one definition for fairness when it comes to AI systems, and context is a key factor in determining what is fair. At Credo AI, we provide tools to our customers to help them determine how fair their AI system is by working with our customers to align on the exact metrics that should be used to assess fairness **based on their use case context**. This work is informed by the industry best practices that the customer's use case is aligned with. Our policy team also focuses on bringing in requirements from regulations, laws, standards, guidelines, and frameworks—and our data science team partners with our customers to understand exactly what their ML models are designed to do, and how they do it; we then create a technical assessment plan designed to evaluate the exact dimensions of the system that are most relevant for understanding whether it is fair in the context it will be deployed.

Given the context-driven nature of AI governance, we advise policymakers to develop context-specific guidance and rules, and transparency reporting will help industry to arrive at the

right standards and rules based on this context - through an iterative process of revealing benchmarks and best practices.

Addressing Risk Now Ensures Leadership in the Long Run

AI is a multi-trillion dollar industry. For the United States to lead in its development, it is crucial to understand the economic and societal outcomes for our nation. RAI is about better, more effective outcomes—it is about producing *more* value for AI builders and consumers by building trust in technology.

RAI is a core competitive differentiator, not just for companies, but for countries. Any government helping to set up RAI requirements on testing and metrics now will have a competitive advantage in first creating and developing accurate methods for assessment and alignment to create trustworthy AI. The work to build trustworthy AI is not *just* about “doing the right thing” and setting “values” that make people feel good. It is about building systems that work better - systems that do not have unintended harmful consequences.

The MIT Sloan Management Review and Boston Consulting Group Report¹ published this month (September 2022) reported that, “RAI Leaders can realize measurable business benefits from their RAI efforts...[which] include better products and services, improved brand differentiation, accelerated innovation, enhanced recruiting and retention, increased customer loyalty, and improved long-term profitability, as well as a better sense of preparedness for emerging regulations. RAI leaders are nearly three times as likely to realize business benefits from their organizations’ RAI initiatives than non-RAI leaders.” This is just one illustration of how investing in trustworthy AI pays off.

When we consider what the effect of algorithmic bias can be on economic contributions to society, we should look at real-world examples, such as an AI system that was deployed in the market and automatically granted lower lines of credit to women than to men. The biased AI system allocated differential lines of credit to a husband and wife with the same address and joint home income, with the woman being granted a much lower line of credit by the AI system than the man. In this scenario, there is no reason a wife should have less credit than her husband, and the system’s decreased accuracy resulted in a loss of economic contributions that women bring to society - an outcome that is unfairly impacts the individuals and is also bad for the business

Another example of algorithmic bias illustrates a loss to the workforce—an AI system that was trained mainly on men’s resumes deprioritized the word “Women’s” when it appeared in resume

¹ Elizabeth M. Renieris, David Kiron, and Steven Mills, “To Be a Responsible AI Leader, Focus on Being Responsible,” MIT Sloan Management Review and Boston Consulting Group, September 2022.

search results. As a result, the AI system missed out on a stellar talent. This example is not just about hurting people, but about creating a system with failures that will have a negative economic impact on the workforce at large.

If we truly want to be a global leader in AI, then our focus should not be on building the most powerful system the fastest, but rather on building responsible technology and support systems that will serve us best in the long run. We will sacrifice the opportunity to lead if we are simply moving quickly for the sake of getting ahead in a way that is not aligned with our societal values.

Conclusion

Credo AI is grateful for the opportunity to appear at today's hearing, and we applaud the Subcommittee's focus on how best to empower organizations to create AI with the highest ethical standards in order to deliver Responsible AI at scale.



NAVRINA SINGH

CEO & Founder, Credo AI

<https://www.credo.ai/>

PROFESSIONAL EXPERIENCE

Past: Microsoft & Qualcomm

Committee Member: NAIAC (National AI Advisory Committee)

Current Board Member: Mozilla

Social Media

Twitter: navrinasingh

Linkedin:

<https://www.linkedin.com/in/navrina>

NAVRINA SINGH is a seasoned customer centric, data driven, global technology and product leader with a proven track record of operationalizing strategy, driving innovation, and commercializing growth. Ms. Singh is the CEO and Co-Founder of **Credo AI**. On a mission to empower organizations to deliver trustworthy Artificial Intelligence (AI) at Scale, Credo AI helps organizations to monitor, measure and manage AI introduced risks. Prior to that, Ms. Singh has led multimillion products and businesses in Enterprise SaaS, Artificial Intelligence (AI) and Mobile over the past 20+ years. Navrina is passionate about responsible leadership and inclusive cultures which she believes are foundational to delivering meaningful impact and transformational innovation to the organization and its people.

Prior to her current startup, Ms. Singh was Director/Principal of Product in **Microsoft** Cloud & AI (2017-2019), where she built Natural language based conversational AI products (chatbots, Virtual agents). In addition to Product, she led the monetization business model to deliver enterprise value, operationalize conversational AI platform technology. Navrina joined Microsoft in 2016 as the Director Business Development for Artificial Intelligence responsible for commercial strategy and partnerships to forge new businesses for Microsoft leveraging AI technologies.

Before joining Microsoft, Ms. Singh spent over a decade at **Qualcomm** Incorporated (2004-2016), where she held multiple roles across product management, strategy, and engineering. From 2011-2015, Ms. Singh was the head of Qualcomm Innovation focused on building new products and creating new market opportunities in Artificial Intelligence, Internet of Things and Mobile across its emerging businesses. As an outspoken voice for Inclusion and Diversity, Navrina founded and led the company's first women initiative focused on getting more women to

leadership roles in technology, equal pay initiatives, transparency across hiring diverse talent etc.

Ms. Singh is a Young Global Leader with **World Economic Forum** (WEF), for her work in disruptive technologies and driving diversity & inclusion initiatives at Scale. Ms. Singh was also a member of the WEF Global Future Council on AI and Robotics, exploring how developments in these fields could impact industry, governments, and society in the future. Ms. Singh is currently also working on global strategic initiatives related to the ethical and responsible development and deployment of AI.

Currently Ms. Singh serves as a member **of the National AI Advisory Committee (NAIAC)**, which is tasked with advising the President and the National AI Initiative Office on topics related to the National AI Initiative. This Advisory Committee was **launched in April 2022**.

Ms Singh is also the executive board member of **Mozilla** Foundation and serves on its audit and Trustworthy AI committee, focused on driving its mission of Open Internet via trustworthy Artificial Intelligence. In the past Ms. Singh has served on the board of the University of Wisconsin- Madison College of Electrical Engineering and on the board of Stella Labs (Hera-Labs), a San Diego based women's accelerator. Ms. Singh is a startup advisor and a technology leader published on **Fortune, Business Insider, TechCrunch, Forbes**, and others.

Ms. Singh holds a MS in Electrical & Computer Engineering from the **University of Wisconsin-Madison**, an MBA from the **University of Southern California** and a BS in Electronics & Telecommunications from Pune College of Engineering, India.

Chairwoman STEVENS. Well, thank you.

And at this point, we're going to turn to our first round of questions, and the Chair is going to recognize herself for 5 minutes.

In hearing your testimony, as I reflect on my time pursuing a master's in philosophy of which my parents never understood why I got, but we were asking the ethical question about artificial intelligence that some ask in the theoretical space that can a AI replace human behavior? Can—does AI threaten what we do as people seeking to overtake, you know, the decisions that we make as people?

Today's hearing is a little bit more instructive to the theoretical question. Today's hearing is saying, hey, we have artificial intelligence, and it is being utilized, but how is it being utilized? How is it being implemented? And is it implementing fairly and accurately for the best outcomes for society and for humanity?

So in 2019, NIST developed the strategy for Federal engagement in developing technical standards and tools for artificial intelligence. And, Ms. Tabassi, I'm just wondering if you could touch briefly because your testimony got me thinking on this, what was included in this strategy and why it is important that we have strategies for engaging in the development of technical standards for artificial intelligence. Has NIST's work on AI management framework revealed new or underdeveloped areas for standardization with regard to trustworthy AI systems? And then, because we want to hear from you on that, but then I want to hear from, I guess, Crenshaw, Mr. Crenshaw, about the—you know, how beneficial it is to industry actors for the Federal Government to lay out priorities and standards for critical technologies and artificial intelligence. Are you using these?

But let's start with you, Ms. Tabassi.

Ms. TABASSI. Thank you very much for the question, Chairwoman. Yes, in 2019, we developed a plan for Federal Government engagement in development of technical standards, and it has several recommendations on bolstering research that's really important for development of good, technically solid, scientifically valid standards, but also importance of public-private partnership and coordination across the government on bolstering our engagements in the standard development and importance of international co-operations on development of standards that are technically sound and correct but also reflect our shared democratic values.

Let me also say that it also lists standards that are related and needed for a trustworthy, responsible AI and of course, many of the standards that's happening for information technology and software systems can be related to artificial intelligence and can be used there but also need for other standards for addressing issues such as bias and explainability and trustworthy.

Chairwoman STEVENS. Great. And, Mr. Crenshaw, I mean, are you using these or, I mean, is this helpful to what you were talking about?

Mr. CRENSHAW. The NIST process is incredibly helpful. It is getting the conversation started and providing the guidance that's necessary for industry to look to. It's incredibly important, too, to have buy-in from the affected stakeholder community. And I have to applaud NIST for the work that they have done through their mul-

tiple rounds of comment, their multiple rounds of public engagement and public meetings to really get this right. And I think it's incredibly important, the work they are doing, that there is a set of guidelines for industry to look to. I think, you know, on the domestic level, that that is a guiding light for industry.

I would note, it's also important to remember standards bodies internationally as well. In order for us to maintain our leadership in this front, we need to make sure that we have American interests represented with American businesses and American policy-makers being aware of that. We do know that our competitors are trying to pack those bodies, and we want to make sure that we are represented as well. I think yesterday—

Chairwoman STEVENS. So are you suggesting more investment?

Mr. CRENSHAW. I'm suggesting more participation, so—

Chairwoman STEVENS. Well, we did just reauthorize NIST, but, you know, Dr. Isbell, what I was kind of getting at was the Turing test, which I know you're familiar with. But I don't know if that's really the question now, is it, you know, in terms of improving these outcomes with AI? And maybe this is too philosophical of a question, but is it the Turing test that that we should be focused on or what is the question that we should be focused on with the fair implementation of AI across a multitude of sectors that are determining our economy at grand scale with 5 seconds left?

Dr. ISBELL. There is no question too philosophical. The short answer is, it's not the Turing test. It's about the actual impact and outcomes on real people. And you have to bring those real people in to understand those outcomes.

Chairwoman STEVENS. And with that, I'm going to now recognize Mr. Feenstra, our Ranking Member, for 5 minutes.

Mr. FEENSTRA. Thank you, Chairman Stevens—Chairwoman Stevens, and thank you for those questions. Thank you again for all witnesses. I really enjoyed your testimonies.

You know, there's extensive research going on in my home State and my universities and—concerning AI, how it's being applied now and into the future. Iowa State's AI Institute for Resilient Agriculture is bringing together experts to lay the groundwork for developing AI-driven predictive plant models to increase the resiliency of agriculture. Researchers at the University of Northern Iowa are aiming to use AI to improve healthcare outcomes, increase privacy, online security, and create predictive maintenance systems for our products. And then in the University of Iowa, they're utilizing AI to improve the effectiveness of cancer screenings, as well as the work to identify and address biases in AI and healthcare models. You know, these are just a few examples that are out there, and they're limitless.

And I would just like to say, Dr. Isbell, I'm an academic also, and I teach—or did teach consumer behavior. And when you start looking at consumer behavior, there's a tremendous amount of AI being used, good and bad.

Ms. Tabassi, I understand that AI won't be replacing doctors, all right? I understand that, won't be replacing nurses. But we also have the opportunity to learn about healthcare-related AI and research, as I just mentioned. Fostering trust in AI will be critical

to utilizing applications such as these in the healthcare sector. And this is just one example.

My question to you, if I can flip my page, can you explain how an AI Risk Management Framework will—broadly applied across the different sectors and industries to minimize the negative impacts of AI systems and maximize positive outcomes? You can use any specific sector examples in healthcare if you wish, but I'd like to know more about that.

Ms. TABASSI. Thank you so very much for the question, Ranking Member Feenstra. And all of the examples that you said just show the potential of AI to really change our lives for better. I'm going to use the last example that you brought up, the cancer screening. So if you have a cancer screening tool, first, as mentioned several times, we wanted to make sure that it's accurate, it's working well, but beyond that the accuracy should also be balanced with associated risks and impact that it can have. So the question comes up about the bias or fairness. Does it advantage or disadvantage certain demographics? Beyond that there's questions about the vulnerability and security and resilience of the AI model, we all hear that AI systems are brittle. Can that cause negative consequences? The issue of the privacy, the data that's used to train the models, can we make sure that the privacy is preserved and the training data are not inferred from the models?

And then on top of that is we heard about the explainability also. If the tool comes out and gives, for example, an outcome or prediction that there is a cancer there, that's a very serious message to be carried to the doctor to the patient. So explainability on how the model decides that there's a cancer there, and another level of complexity, the explanation needed for physician versus technician versus patient is different. AI RMF is trying to provide a shared lexicon, interoperable way to address all of these questions, but also provide a measurable process, metrics and methodology to measure them and manage these risks.

Mr. FEENSTRA. Thank you so much for that. That's great information.

Mr. Crenshaw, in your testimony you say that trust is a partnership? I 100 percent agree. And only when government and industry work side by side can trust be built. How did NIST work with industry in developing the AI Risk Management Framework? And how is having a tool like the framework going to strengthen consumer confidence when it comes to building trust in the AI systems?

Mr. CRENSHAW. Well, I think as I said, Congressman, trust is essential. And I think NIST has done a great job of really instilling trust in their work with the business community by being open and transparent. If you look at the the comment record, it's comments from across the board, everyone from civil society all the way to industry and developers. And they're really looking to develop a robust record. That I believe is a really great example for other agencies as they're looking at tackling this issue to look at. So they've had multiple stakeholder sessions. They've come in and actually spoken with our members and tried to get a good feel for where they're at. And it really—the partnership has been excellent, and

I think it's a great example for other agencies moving forward in this space.

Mr. FEENSTRA. Thank you, Mr. Crenshaw, I have questions for Dr. Isbell and Ms. Singh, but I ran out of time. So with that, thank you for your testimony. I yield back.

Chairwoman STEVENS. Great. And with that, we're going to hear from Dr. Foster for 5 minutes of questioning.

Mr. FOSTER. Thank you, Madam Chair.

So my first general question is this discussion converging? You know, I've been chairing the Task Force on AI and Financial Services for the last several years, and it strikes me that the complexity of AI behavior is increasing much more rapidly than our ability to categorize and regulate it. You know, an example of that is a simple neural net classifier that's operating on a static data set to calculate credit scores or something like that has a relatively—it's an enormous, but it's a relatively finite range of behaviors to categorize, OK?

On the other hand, interactive AI, which is an agent which is learning from other intelligent agents and guiding its behavior, has an enormously larger space of behaviors to characterize. And I just don't even see how you can possibly explain how an intelligent agent might react in any given circumstances. Like you can say general things like, you know, this child is a fast learner but makes a lot of mistakes, but that doesn't give you the granularity of detail you need.

And so I'm just wondering, since you've been all thinking about this, do you get the feeling that it is converging or not? No? Dr. Isbell?

Dr. ISBELL. The short answer is no. The problems that we're talking about are exponential. All of our solutions are linear. You might as well ask the question whether human behavior is converging and we know how to understand or regulate that. And of course, the answer is no, but that does not mean that there are not things that we can do to make progress. And I do think a lot of the discussions that we've had just in the last couple of years around fairness, accountability, thinking about how to educate people to be in the—to be a part of these discussions do make real progress, and that progress doesn't—is very sudden, and makes very sudden changes, so it's a good thing.

Mr. FOSTER. Any other thoughts on this? Yes, Dr. Singh?

Ms. SINGH. Congressman, I think that's a great question. I believe we are making progress toward convergence. But one of the key areas that I spoke about earlier is how important context is to this work. So one of the core acts that we have as standards emerge in this space is really thinking about context, the applications, and how we can make progress toward the right metrics and assessments, along with the specific reporting requirements. And we are seeing globally as well as the great work that NIST is doing that there is a convergence that has started to happen in terms of having those contextual conversations.

Mr. FOSTER. Any other thoughts? It's a huge question. Let's see—many of you have emphasized education and the need for an educated public. So if you had to choose between a public that knew statistics or knew calculus, which would you take? I'm a physicist,

so I naturally lean toward calculus, but it seems like what I use every day as a politician, statistics are relevant. And probably for AI, I think you're in the same bin. And do you have any—well, all right, Dr. Isbell—but you have to deal with curricula, so you're on the seat again.

Dr. ISBELL. I'm not speaking for all of my colleagues. I think the answer is, if I had to choose for most people, it would be statistics, but I'd also like them to know information theory and linear algebra. But fundamentally, it's about problem solving around data mattering as opposed to just the algorithms and the processes that you go through. And with that you can solve a lot of the problems or at least address and think about the problems that are coming down the pike.

Mr. FOSTER. Any other thoughts from any of you? What do you use every day, statistics or calculus? I think—yes, machine learning. It's—backpropagation is the chain rule, and I don't think there's much other calculus anywhere in it. But anyway, the—now, actually, this was for Mr. Crenshaw. You've emphasized international competition, and it strikes me that a lot of the countries that are clobbering us, you can't get out of high school without knowing calculus and probably statistics. There's all sorts of people showing up at school boards, you know, unhappy that we're not supporting their preferred theology or mythology. But very few school boards are being inundated by people, you know, demanding that our kids know statistics and calculus. What—is there some—is there work to be done there?

Mr. CRENSHAW. There's definitely work to be done on the education front. We need to prioritize STEM education to ensure that we have the fundamental knowledge base for students across the country to get into this field because we are going to need more coders and ethicists in this field who actually can assist with our leadership.

The other thing I think would be important to note, too, is that we also need to make sure that we have talent in this country and retain talent and still attract talent. And one of the things that we found out through our AI Commission is that we, you know are going to lose the talent race if we don't deal with our immigration issues in this country as well and make sure that we can retain talent after we've educated them here in the United States, make sure that we can keep our talent to ensure that we have people who know how to make ethical AI work.

Mr. FOSTER. Thank you. And we in a bipartisan way on this Committee have been doing everything we can to try to drag that across the finish line. I think we came within one Senator of doing something significant in the *CHIPS and Science*.

Anyway, my time's up and will yield back.

Chairwoman STEVENS. And with that, we will hear from the Ranking Member of the Full Committee who we're so grateful is here, Mr. Lucas for 5 minutes of questioning.

Mr. LUCAS. Thank you, Madam Chairman. Ms. Tabassi, in the AI Initiative that we passed in Congress last year, we gave NIST the difficult task of defining what makes AI safe and trustworthy. Can you walk us through the process of how NIST determined that definition of trustworthiness? And while you're thinking about that, do

you think this measure of trustworthiness also helps with the measuring of fairness in AI systems, please?

Ms. TABASSI. Thank you so very much, Ranking Member Lucas, for the question. In terms of the process of developing a definition of the trustworthiness, I want to thank the kind of work that has been mentioned about the NIST process. But the process has been an open, transparent, collaborative process. There has been many definitions and proposals for definition for trustworthiness, so we ran a stakeholder-driven effort to converge to the extent possible on the definition of the trustworthiness. And that, as was mentioned, include rounds of workshops and public comment and a listening session. So that was the process.

Your second part of the question is about the fairness. So fairness is one of the aspects of the trustworthiness as it's mentioned in the AI RMF. And fairness, as it was mentioned, is a complicated concept because it can depend on societal values and can change from context to context. But that's also part of one of the aspects of the trustworthiness mentioned in the AI RMF.

Mr. LUCAS. Ms. Singh, in your testimony, you illustrate why you cannot have a one-size-fits-all definition of an algorithmic fairness. How does the AI Risk Management Framework exemplify this?

Ms. SINGH. As I previously stated, I really commend NIST for the Risk Management Framework and how they're thinking through not only mapping different applications, but measuring and then overall management of those. At Credo AI, we are really focused on operationalizing responsible AI tenets and ensuring that continuous oversight and governance is provided of these systems. And I think for us it is really critical that there are governance assets based on the context of AI application that gets generated that inspires that trust that Ms. Tabassi was just talking about.

Mr. LUCAS. Mr. Crenshaw, do you foresee U.S. industry widely adopting and utilizing the Risk Management Framework since it's a voluntary tool, or will it need to be incentivized? While you're thinking about that, do you anticipate U.S. standard bodies will play a role in encouraging the utilization of the framework?

Mr. CRENSHAW. I think there's definitely a role there. I think they also have really gotten the conversation out about the need to develop standards. When it comes to the NIST Risk Management Framework, I think what we've seen of it is promising. Obviously, we'll have to comment on the final product when it comes out. But I think it is a promising product. And, you know, I think, given the fact that we've had such robust stakeholder input, I do anticipate that, you know, given the direction things are going, we definitely could see stakeholder engagement to support the framework. And I think that's a good thing because we need guidelines and standards to get behind so we can develop trust.

Mr. LUCAS. Ms. Singh, do you have any thoughts on this point?

Ms. SINGH. I think multistakeholder engagement is going to be critical in the process. And as—you know, we've been invited to give feedback on the NIST RMF, and we've done that actively over the past couple of months. As mentioned, I think there's a little bit more work to be done in terms of ensuring that we are looking at different applications and context.

Mr. LUCAS. Ms. Tabassi, any thoughts?

Ms. TABASSI. In terms of the adoption, I think that the adoption and use of the AI RMF would be based on the value that it provides and also giving awareness that these things exist is also very important. I thank again the Committee and all of my panelists for the kind words about the process. And in terms of the context and specific use, agreed that a lot more work needs to be done. And we have a call for contribution particularly for that.

Mr. LUCAS. One last question, and I come back to you, Ms. Tabassi. Why is it important for democratic nations to lead the development of international standards for trustworthy AI systems?

Ms. TABASSI. I believe it's important to affirm our shared democratic values of openness, protection of democracy and human rights, and design and develop technologies that operationalizes those values. And we need standards for technologies that are rights-affirming and show those values.

Mr. LUCAS. Just the way I intend to answer questions about that in my town meeting someday. Thank you. Yield back, Madam Chair.

Chairwoman STEVENS. With that, we are going to hear from the Congresswoman from North Carolina, Ms. Ross, for 5 minutes of questioning.

Ms. ROSS. Thank you very much, Chairwoman Stevens and Ranking Member Feenstra. And thank you to the panelists for joining us today. On April 29th of last year [inaudible] represents a larger problem of cybersecurity and privacy issues in this country. AI innovation happens fast, and we need legislation that's equipped to grow into this quickly expanding sector. For my constituents in the Research Triangle and for national security more broadly, we need to invest in long-term structural infrastructure that ensures better cybersecurity and privacy in our tech sector. We also need to look at how AI affects the arts and our creators, and we all have many of them in our district. So I look forward to hearing from our witnesses on how we can ensure that systems of machine learning can be created with consideration for individual privacy, corporate privacy, intellectual property, and national security. But since none of the folks who have asked questions yet have talked about intellectual property, and I serve on the Judiciary Subcommittee on that, I'm going to ask Ms. Tabassi—I'm sorry if I mispronounced your name—to say I want to thank you for your important work on the draft of the Artificial Intelligence Risk Management Framework.

But I also want to talk a little bit about intellectual property because the United States takes our intellectual property protections very seriously. And without those protections, there's a significant threat to American creativity, ingenuity, jobs, and our economy. And AI offers opportunities to artists and creators to enhance the creation process in many ways, but that also presents risks. And there are services and sites available today that use art, books, music, and other American-made works as inputs to train AI.

Based on what is happening with image-generating AI currently on the web, we can already see that artists will have to compete with AI creations in their own style and trained on their own content when they were either—neither consulted nor compensated for this. And as a matter of fact, there was a recent article that I just

read about that. Is this issue on NIST's radar screen, and what can we do about it?

Ms. TABASSI. Thank you so very much for the question, Congresswoman. And we have actually received comments to that effect to AI RMF. And that's a serious problem, certainly something that would be part of the discussions in the future drafts of the RMF. A lot of work needs to be done, and that would definitely be part of the discussion. Thank you.

Ms. ROSS. OK. I do have a couple of other questions. Dr. Isbell, your written testimony talks about the Marshall Project and the use of risk assessment in the criminal justice system. How can transparency increase the ability of individuals to protect their information and avoid undue scrutiny? And to whom should individuals direct their concerns if they believe that their data has been misused?

Dr. ISBELL. So it's a very—it's actually quite a difficult problem because the data that we have is out there everywhere, and we leave a trail everywhere that we go. Fundamentally, there has to be policy and there has to be infrastructure. This is a role that government has to provide a mechanism by which people can deal with issues where their data had been misused. It is not a thing that will naturally come from industry. It is not a thing that naturally comes from the educational sector. It is something that has to be dealt with by the legal system.

Ms. ROSS. And can you tell us about any law enforcement practices that we should be aware of as we're considering changes to the legal system?

Dr. ISBELL. Well, I think the short answer is you have to think very carefully about and look at the way that the systems that are out there are currently being used and how they're currently being misused. And having done that, it takes you down a path toward understanding how you have to try to address those one at a time. It's a pervasive thing that touches everything. I—we don't have time to talk about this now, but you—earlier, someone made a comment that doctors will not be replaced by AI. Well, they're already being replaced by AI, and they're being done in an unregulated way that's having an impact on people. And you have to be—you have to recognize that and you have to address it context by context and one case at a time.

Ms. ROSS. Thank you, Madam Chairman, and I yield back.

Chairwoman STEVENS. Great. And with that, we're going to hear from Dr. Baird of Indiana for 5 minutes of questioning.

Mr. BAIRD. Thank you, Madam Chair. And I appreciate you and Ranking Member Feenstra for holding this important hearing. And I really appreciate, I always do, the expertise of the witnesses and their ability to answer our questions and it's very important and very specific.

My first question goes to Dr. Isbell. And I want to know what role have universities played in the development of the AI Risk Management Framework? And more broadly, how are universities helping to shape the future of AI by engaging in public-private partnerships, Dr. Isbell?

Dr. ISBELL. So the—higher education in general is—universities have participated by being invited in and being a part of the con-

versations. Individuals and organizations have continued to participate in all of these discussions around standards, including things that NIST has done, but also through operations of institutes that have been created, for example, by NSF. What the universities do, what our role is, is to do the basic research that exists to create the basic research, ask the basic questions, and then educate the students who are going to go forward and to do that work. A lot of the work that we do, a lot of where we play that role isn't actually identifying the fundamental problems. That is sort of what academic freedom allows you to do, and that's what we continue to do. The environment that we create is one that is—that allows us to ask these questions and to make them available for industry, to make them available for government to take the next step. That's what we do.

Mr. BAIRD. Well, thank you very much. Ms. Tabassi, to your knowledge, has the People's Republic of China developed a similar tool to the AI Risk Management Framework? And what about any of our allies? And so what role if any has NIST played in sharing findings and the best practices with the international community, particularly our allies? So if you have any thoughts in that area, I would appreciate it.

Ms. TABASSI. Thank you so very much for the question, Congressman. In terms of cooperation and collaboration with our allies, the stakeholder engagement effort that we run includes our international partners, so they have been involved in terms of providing input to the AI RMF, coming to our workshops and participating in those events, but we also interact with them and talk with them in forums such as Trade and Technology Council, QUAD, or OECD. So there is a good, strong, robust engagement going on that way.

Mr. BAIRD. Thank you. Then my last question goes to Ms. Singh. So in creating the tools to help companies develop responsible AI, what are some of the most common concerns with AI systems that your company has seen?

Ms. SINGH. Thank you so much for that question. You know, if responsibly and not built artificial intelligence is going to have very varying impacts on different use cases. So across the companies that we work with, one of the things that is critical is, again, really having a holistic view of from the time you're designing the AI system to the actual use, making sure that you're interrogating the technical systems, you're interrogating the processes, as well as you're interrogating the outputs. So this goes back to really identifying any unintended consequences that could appear in the entire AI lifecycle.

Mr. BAIRD. Thank you very much. And I appreciate the witnesses' responses. And with that, Madam Chair, I yield back.

Mr. MCNERNEY [presiding]. Well, I was going to—I think I'm the next questioner, and I was going to thank the Chairwoman for this great hearing, but I certainly want to thank the panelists. Your testimony is great. What a great, incredible subject. I want to get right to questions though.

Ms. Tabassi, how might standards and assessments be developed and—for explainability and interoperability

Ms. TABASSI. We do that the same way that we do for any type of other standards. With true stakeholder engagements and work-

ing with a whole community. Broad stakeholder engagement underlines everything we do at NIST and explainability, interoperability are difficult, complex topics. We do have some foundational research going on. Our researchers are working on this, but we also augment it with the work of the whole community.

Mr. MCNERNEY. OK. Well, I've been on standards committees, and I know what kind of work goes on. So you're saying it's a similar process or would be a similar process?

Ms. TABASSI. Correct. Part of it, doing the internal research, providing technical contributions, working with the whole community on strengthening the research and taking the contributions to the standard development organizations and hopefully see them through become international standards.

Mr. MCNERNEY. Thank you.

Dr. Isbell, in math and physics, systems and solutions are considered unstable if small changes in the initial conditions result in large changes in the solutions and outputs. Are AI systems unstable in terms of the data input? And, if so, how can that be mitigated?

Dr. ISBELL. Some of them are. There's a wide range of ways of doing AI and machine learning. Some of them are quite stable, and some of them are less stable. There's a lot of theory behind this and a lot of work that's been done over decades to get there.

I think the most important thing actually is not the sort of instability that you're talking about with small changes but that we don't actually understand how the set of parameters that go into the way that we build these systems have that impact. It's actually less about the data in that sense and more about the way that we build the systems in the first place. And that has remained largely unexplored.

Mr. MCNERNEY. Well, thank you. That'd be a great area for research. Thank you.

Ms. Tabassi, can you touch briefly on what's included in the strategy of engaging technical standards for tools for artificial intelligence?

Ms. TABASSI. Thank you for that question, Congressman. And, yes, happy to. So that strategy for working toward the standard was developed in 2019. And we are basically implementing the recommendations of that plan since it has been developed in 2019. What's in the plan? Basically talks about standards, standard development processes, talks about AI standards, what's needed, and concludes with recommendations on what's needed to maintain U.S. leadership in development of the technical standards and recommendations very broadly is about strengthening research for development of scientifically valid standards, public-private partnership, to be able to do that research and build those foundations, and international cooperations for development of standards.

I just also want to note, that plan was also developed in a stakeholder-driven effort with a lot of input from the community.

Mr. MCNERNEY. Thank you. So what what extent is the United States already collaborating with the EU and other likeminded nations on developing standards for trustworthy AI?

Ms. TABASSI. Multiple ways. One of them is by expert-to-expert scientists working on what we call pre-standardization research to

actually provide the scientific foundations for the standards and then cooperation by to the standard meeting and seeing them through to become international standard, but also at the forum such as TTC and QUAD.

Mr. MCNERNEY. Well, thank you.

Mr. Crenshaw, I didn't want to leave you out. Would the Chamber and presumably many U.S. businesses support the development of a United States AI regulatory law?

Mr. CRENSHAW. I think, given the state of the technology, we believe it's premature to get into prescriptive regulation. We support voluntary frameworks like we see at NIST. A few areas, though, I think, you know, we would like to see regulation is for things like consumer privacy. We'd like to see a national standard put in place. But at the same time, we want to make sure that the process at NIST can work itself out first before we start making any kind of determinations on regulation. And it's also an issue, though, our own AI Commission is working through as well to make recommendations for.

Mr. MCNERNEY. Thank you. My time has expired, and I'm going to call on Mr. LaTurner. You're up for 5 minutes.

Mr. LATURNER. Thank you, Mr. Chairman. I appreciate it. Ms. Singh, in your testimony, you talk about the need for policymakers to establish benchmarks for fairness when it comes to responsible AI, yet you also talked about how industry-specific and context-driven artificial intelligence factors preclude standard-setting bodies from creating a one-size-fits-all metrics. In a context-specific field, how can Congress create meaningful regulation that ensures AI systems retain algorithmic fairness?

Ms. SINGH. Thank you so much for that question. I think the work that NIST is doing is a good example of the public-private partnership that is needed to ensure that we are doing thoughtful policymaking and standards that are very context-specific. As I've stated previously, you know, in artificial intelligence, the question that we should be asking ourselves right now is how can governance and oversight keep up with the development of artificial intelligence? And so we believe that standards are going to be critical, especially as we think about transparency reporting. And transparency reporting, is going to be a complete view into the AI lifecycle that can help with benchmarking.

Mr. LATURNER. What could we be doing differently with our— with Congress and the public-private partnerships? Do you have any recommendations on how we could be doing it better?

Ms. SINGH. Yes, thank you so much for that question. You know, we've given some feedback to NIST on that. I think we have to really step back and think about the AI application, as well as what the impact to the stakeholders within that AI application is. And I think going back to context-centric metrics, as well as context-centric reporting requirements is one of the first steps we believe is going to help move this industry forward.

Mr. LATURNER. How can developing responsible AI give the United States an economic and societal competitive advantage over other countries

Ms. SINGH. Thank you. I think that is a fantastic question. We at Credo AI believe that responsible AI is a competitive advantage

because it is not only going to help United States and the companies here deploy AI with confidence, but as we make sure that the standards that emerge which are aligned with our societal values, that is going to promote more consumer trust, which, as you can imagine, is going to further bolster our leadership in artificial intelligence.

Mr. LATURNER. Thank you, Ms. Singh.

Dr. Isbell, you state in your testimony that there are many occasions where tech workers cannot be certain how AI algorithms reach the correct answer, and these algorithms are known as, quote, black-box models. If for any reason these types of algorithms reach an incorrect or biased outcome like the ones you describe in your testimony, it can be nearly impossible to diagnose. If we want to solve the problem of black-box models by making an algorithm's data set more transparent, then what countermeasures can we take to bolster AI security from hackers? To your knowledge, are there any examples of AI developers that have already—that are already addressing this issue?

Dr. ISBELL. So there's a great amount—there's a large amount of work that's being done in academia at the level of basic research to understand differential privacy, to understand how it is that people can interfere and break into the way that machine learning algorithms actually work. So there's a lot of work. It's in early stages, but a lot of great stuff is being done. How much of the—not a lot of that has necessarily been deployed in the systems that are out there now I think in large part because the incentives haven't necessarily been there.

What drives industry and drives the people who build these systems and deploy them to do—to touch on this is requirements that either through the market or through policy, that if they don't do this, they're simply not going to be able to deploy their systems and to have them used and adopted by large groups of people.

So there's a lot of work that's been done out there, a lot of specific things. I would start with differential privacy, and there's lots of researchers that have done great work on this. But at the end of the day, it's really going to be about creating the incentives for people to want to take advantage of what we know in order to keep things secure.

Mr. LATURNER. Thank you. Mr. Chairman, I yield back.

Chairwoman STEVENS. Great. And with that, we're going to hear from Mr. Beyer of the Commonwealth of Virginia for 5 minutes of questioning.

Mr. BEYER. Thank you, Madam Chair, very much. And thank the witnesses for really interesting feedback. But also thank my colleagues, Democrats and Republicans, for some very good questions.

Ms. Tabassi, I know you take on this tremendous task of managing, developing the AI Risk Management Framework. You heard from Mr. Crenshaw what the Chamber is doing with its commission. And I think you've heard pushback about how we're not ready to have mandatory standards, that we're still so early that we're—we don't want to overreact. We don't want to overregulate. But at the same time is it not naive to think that we can make this voluntary indefinitely, that at some point there won't be a need for

clarity in terms of what is demanded and expected from businesses in AI?

Ms. TABASSI. Thank you very much for that very thoughtful question, Congressman. So NIST AI RMF is a voluntary framework just like any other frameworks that NIST has developed. And the use and adoption of that, at least, I believe, would be based on the value that it provides. And another strength of the voluntary process that we are doing is based on the stakeholder engagement and stakeholder-driven process that we are following in development of this voluntary tool. It gives the opportunity to the whole community to provide their input, their comments. So by the end, the final tool would be a more effective resource that everybody that participate in development of that would have a buy-in in that.

So by that, I think, having the value on using this and having buy-in because of participation in the process of developing it, would help with its adoption. NIST is a nonregulatory agency, and the things we put out are voluntary.

Mr. BEYER. We know that, so thank you. I understand you're nonregulatory and ultimately it will come back to us and then come back to us just based on dangers.

Dr. Isbell, I was fascinated by your testimony. Because so much of what we talked about today is concern about biases, but you also had a wonderful paragraph about the upside of machine learning and artificial intelligence. Can you expand on that a little bit? It seems to me that we as human beings dramatically underestimate the potential for what artificial intelligence can bring humanity.

Dr. ISBELL. So there's a particular law, and I forget what—escapes me right now. But what the law says is that we overestimate the short term and we underestimate the long term. And I think that's exactly what's been happening with AI. There was a lot of hype back in the 1970's and 1980's before the AI winter with all the great changes that AI was going to bring to the world. They were wrong. They were overhyped.

But it's turned out that the impact that AI has had has been profound and far deeper than anything anyone even imagined back then. It has infiltrated every part of our life, and I use infiltrate in a positive way. We will be doing a better job of detecting when people are sick in ways that we were never able to do. We will be able to help people to make decisions they otherwise would not have ever been able to make. We will be able to connect with one another in ways that we have not been able to connect with one another before. And a large part of it will be because of computing, and it'll be because of AI. It's all very positive. The opportunities in front of us are huge, and it will take us—it will help us to solve big problems that we currently have a hard time thinking through and those problems over decades and even over centuries.

The problem that we have, of course, is that we have to set up the incentives to allow people to do that, and we have to make certain that everyday people understand enough of what's actually going on so that they can make rational decisions about how to use that technology in their own lives.

Mr. BEYER. Dr. Isbell, I'd love to have a question for the record if you could find one of your research assistants to find out the name of that law.

Dr. ISBELL. I will.

Mr. BEYER. Dr. Vint Cerf told it to me 30 years ago, and I've always attributed it to him, but it probably has a deeper root.

Dr. ISBELL. Absolutely.

Mr. BEYER. Very powerful.

Dr. Singh, one quick question. You know, we've been struggling with facial recognition technology on police bodycams. Now, is this something that you're working on, too, that the notion that people of color, especially women of color, are picked up inaccurately much more frequently than others?

Ms. SINGH. Thank you so much for that question. We at Credo AI work across a diverse range of applications, including facial recognition. And as I stated previously, I think any artificial intelligence that is not developed responsibly is going to impact all of us, and especially the marginalized communities, which in the past have been excluded because of gender, ethnicity, color, are at a higher disadvantage here. So building responsible AI is not just competitive advantage, but it is going to serve humanity really well.

Mr. BEYER. Madam Chair, I yield back.

Chairwoman STEVENS. Thank you. And with that, we're going to hear from Mr. Gonzalez of Ohio for 5 minutes of questioning.

Mr. GONZALEZ. Thank you, Chairwoman Stevens, Ranking Member Feenstra, for holding this hearing. Thanks to all the witnesses for your testimonies.

Ms. Tabassi, we talked a little bit about the AI Risk Management Framework, and that was helpful. I'm curious, has China developed a similar tool? What is China doing specifically around this?

Ms. TABASSI. Right. So I believe it was in 2017 that China put a very ambitious domestic AI plan out. To the best of my knowledge, there isn't anything that they're doing similar to the AI RMF. If they're doing it domestically, I don't know. But—yes.

Mr. GONZALEZ. OK. Thank you.

Mr. Crenshaw, I'm going to switch to you for a second. Unlike most countries that have a top-down, government-led approach, the United States has a bottoms-up, industry-led approach to standards setting, which I think is appropriate. We employ a voluntary system which relies on industry participation and leadership. This market-driven approach enables competition, ensures transparency, and takes advantage of consensus-building to drive us to the best possible outcomes. Can you explain how the U.S. approach to AI through the AI Risk Management Framework drives innovation?

Mr. CRENSHAW. Well, I think it's interesting to know, during one of our hearings, we actually had one of the cochairs of the National AI Advisory Committee come testify, Miriam Vogel. And she said the reason we needed to maintain leadership in this country is because we have a brand of trust compared to other countries. And it's important that we have standards in place that are voluntary, that will be adaptable to this new and developing technology but at the same time will look at things like risk. And it's important that we have real firm guidance in place.

And another—I think, as I said before as well, when it comes to international standards bodies, we need to make sure that the

United States is well-represented. The *CHIPS and Science Act* actually helped provide funding to ensure we can participate in that space. But, you know, at the same time, too, as companies look at things like developing implementation for compliance or following guidelines, if they go out there and say we're following this guideline and then they're found not to be, there is some teeth there.

Mr. GONZALEZ. Yes.

Mr. CRENSHAW. So there are agencies that can enforce there as well.

Mr. GONZALEZ. Great.

Mr. CRENSHAW. So there is great trust to be had by establishing leadership and trust against other countries.

Mr. GONZALEZ. Dr. Isbell, with your role on campus as a Professor and Dean, what do you believe the appropriate role of the university is—are in shaping the future of AI?

Dr. ISBELL. Twofold. One is to do research. We have one of the best systems in the world around basic research. Our research ones are amazing. And all the way down to our research twos and even our community colleges are able to bring people in and to think about and engage in the conversation around AI or any other large, important issue. So the research is important, and maintaining and supporting that is important.

But the second and perhaps the most obvious is the fundamental mission, which is educating people, not just educating the people who are going to do the research, but I think importantly, and especially when it comes to AI and machine learning, is educating everyone else who is not going to do AI and machine learning research but will be affected by it, who will be adjacent to it, and will be far away. As I told my son who's deeply into history, you will not be able to get a degree in history in 5 years without knowing machine learning and AI because it's still going to be data-driven. And so our responsibility is to make certain that everyone is a part of that conversation.

Mr. GONZALEZ. Great. And then I agree 100 percent on the research point, actually, on both points. But, you know, one thing we talk about a lot on this Committee is how do we get the research—the incredible research that's happening on our university campuses out into the public space and then driving innovation in the private sector? So what do you think we need to be doing to have a—I'll just call it a more robust sort of flywheel of research taking place on college campuses, leads to innovation, leads to private companies, et cetera, et cetera?

Dr. ISBELL. So we actually do pretty well with that, I think, but I think the biggest problem right now is that there's a mismatch between what the company—pick whatever your favorite company is—wants to do in the next 6 months to a year versus what the basic research that's looking out 5 or 10 years actually is. Support through organizations like NSF, for example, to help partner with those companies, to partner with industry to help do the basic research, universities, I think, is the best way to get that translational work done from the lab out into the world. And when it works, it works very well.

Mr. GONZALEZ. Thank you. I yield back.

Chairwoman STEVENS. Thank you.

With that, we'll hear from Congressman Sherman of California for 5 minutes of questioning.

Mr. SHERMAN. Thank you, and thank you for allowing me to participate in this Subcommittee's hearing. Without objection, I'd like to enter into the record an article I wrote 22 years ago, "Engineered Intelligence: Creating Our Successors' Species."

My line of questioning is going to be about things that won't affect us until the second half of this century. But since they relate to whether humankind will continue to be in domination of the planet Earth, they're important. We're—right now, the computer engineers and the bioengineers are racing to create a new level of intelligence. And the last time there was a higher level, a new level of intelligence appeared on the planet is when our ancestors said hello to Neanderthal. It did not work out well for Neanderthal.

So my focus is on whether we're going to see artificial intelligence that has general intelligence, self-awareness, and what I call the ambition, or survival instinct, or care. And that third thing I should go into more, I tend to think that our successor species would be biological because even the dumbest worm seems to care if you try to turn it off or kill it, whereas the smartest computers we have so far don't care if you unplug them.

So my concern is what are we doing to prevent or monitor for general intelligence, self-awareness, and ambition or survival instinct? Or are we just going to ignore those issues and focus on things that affect us in the next decade? Ms. Tabassi?

Ms. TABASSI. Thank you very much, Congressman, for the question. It's hard to determine when or if we can reach or the community can reach to an artificial general intelligence. I will say that that's—

Mr. SHERMAN. Well, I think we're going to get there someday.

Ms. TABASSI. Right.

Mr. SHERMAN. We just don't know—

Ms. TABASSI. Very good, very good. So we don't know when we're going to get there. So from the NIST point of view, we think that that's one reason to work on foundational principles. That's why it's now timely—

Mr. SHERMAN. Is anybody doing any technical research about how we can get very useful computers, that we somehow put something in there, a governor if you will, that prevents general intelligence or prevents self-awareness, or prevents ambition and caring? Is anybody doing the research as to how we can get what we want without getting what we don't want?

Ms. TABASSI. I'm not aware of that research being done at our laboratory at NIST, across the academia, and the community. I don't know. Thank you for the question.

Mr. SHERMAN. I'll ask the other witnesses. Is anybody aware of us trying to prevent, as we try to harvest the benefits of artificial intelligence, the creation of an ambitious, self-aware computer that may very well decide that we're irrelevant to this planet? Is anybody figuring out how to do that, or is it just an issue we're all aware of but aren't really trying to confront? Does anyone just—yes, Mr.—yes, Doctor?

Dr. ISBELL. So I guess the—yes, and thank you for the question. Actually, you know, one of the reasons I got into AI in the first

place were these what I'd consider pretty existential and philosophical questions around what does it mean to build intelligence? I think the answer is that people discuss these issues all the time. They try to figure it out, they try to work it through. We don't have any large research, at least that I'm aware of, any large research agendas around preventing the issue—preventing general intelligence in part because we have no idea how to get there from here. And I think one of the things that I would leave——

Mr. SHERMAN. What about those two other issues, how to prevent self-awareness, how to monitor for self-awareness, how to prevent ambition or survival instinct, how to monitor for survival instinct?

Dr. ISBELL. I don't think it's done in those terms. I don't think it's done in those terms. It's done in simpler terms around preventing harm.

Mr. SHERMAN. Well, we're going to concentrate on the harm that could occur in the next decade——

Dr. ISBELL. That's right.

Mr. SHERMAN [continuing]. The Nation or artists that lose their creativity and the benefits of their creativity, and it doesn't seem like anybody's worried about the problems we'll confront in the second half of this century. And with that, I yield back.

Chairwoman STEVENS. Great. And with that, we're going to go to another round of questions because we're just having so much fun here. And the Chair is going to recognize herself for 5 minutes. I think this question about where and how we're determining the ethics is very important. Obviously, we have so much respect for NIST and an understanding of the role that standards play. We could go philosophical again and ask our standards, ethics, and how the ethics arrive out of standards that come from rigorous processes that are inputted by—you know, we talked about the companies, we've heard from Dr. Isbell about the people, the people element that needs to get involved with the standards.

But, Dr. Isbell, some universities are already including ethics as a curriculum and long have. You go into a philosophy department, you're going to get an ethics course. Hopefully, people take it. But ethics as a curriculum requirement for computer science degrees in particular, a great start, but it's often obviously sometimes a separate course and may not be directly connected to what students are learning in other courses.

You've changed your approach at Georgia Tech, and so I just wondering if you could elaborate on what you're doing to integrate ethics education and how you're assessing its effectiveness. And I also just—because that's a question I know you can answer it, but I just really want to applaud you for a segment in your testimony that I encourage everyone to look at where you said computing has long been an intellectual wild west where things change so fast that the priority was always to fix—to find what's next, to find the better solution. Now, we've succeeded in finding solutions so good that they are intertwined in nearly every area of our personal lives and communities. So can our laws move fast enough? Can our ethics move fast enough? And where and how do we find this arising? Thank you.

Dr. ISBELL. Sure. Thank you for the question. I really appreciate it. I will say that, you know, people in my field have spent 40, 50 years trying to convince everyone that what we did was really important, and it turns out, we were right. And then what we're living with now are the consequences of having been right.

So when it comes to ethics and responsibility, I think the—you know, Georgia Tech, we've had that as a requirement for CS going back at least about 30 years. But what we had done wrong—and not just us, but I think the way that we approached this—is that we treat it, as you say, a separate class, something that gets stapled on at the end. It's a requirement. Nobody takes it till their last semester. It doesn't get integrated into the rest of the curriculum and it can't.

So one of the things that we did recently is we kept it as a requirement, and we made it a prerequisite for our junior yearlong design classes. So by the time you're a sophomore, you know just enough to be dangerous. You're at a place where you're being forced to think carefully about the consequences of the systems that you build, and then you're asked to build such a big system. This is before you take Intro to AI. This before you take Intro to Machine Learning. This is before you take Introduction to Cybersecurity and Privacy. So it puts you in a place where the people further down the chain can actually now ask you the direct questions that they couldn't do before because you wouldn't have the language or the experience to be able to do that.

That is what's important. When we claim that something is important, we have to operationalize it in our curriculum in the way that we teach people from the very beginning and not toward the end, which is the natural thing to do if you aren't very careful about how important you think that it is.

Chairwoman STEVENS. And certainly to Mr. Crenshaw, I'm sure you have some thoughts about this as well. And, you know, we applaud the the point about, hey, we want to drive a—you know, American leadership of what we're doing with artificial intelligence.

And thank you, Ms. Singh, by the way. I've just so thoroughly enjoyed your—not only your testimony, but the answers to your questions. But how do we balance these things out, right? You know, we sometimes see, you know, too much of a good thing, per se. And we don't—you know, we like standards. We're doing standards. You've said you like the risk management. But, you know, in some ways, right, we see companies getting pushback because they haven't self-regulated and the ethics component isn't there. And so, you know, where and how do we find that balance? And maybe that's articulated through boards. Which—how does that populate? And maybe Ms. Singh can chime in, too.

Mr. CRENSHAW. I think it's critically important, one note to make, that we have the critical decisionmakers in companies involved in this process as well. Not only do technologists have a role, but C-suite does as well. And also, you know, we need more education out there about the need to build in ethical AI into standards for companies and how they operate. I've talked to some companies that are actually developing their own ethical frameworks and have full-time ethicists who are being brought on. We had a

hearing actually at the Cleveland Clinic about 4 months ago in which they've now brought on an ethicist as well, as they're using AI to treat their patients. So it's important, and I think companies are beginning to see this.

Chairwoman STEVENS. Yes.

Ms. SINGH. Thank you, Chairwoman. I think, today, we've established that AI is not a technical problem. It's a sociotechnical problem that really needs multistakeholder perspective and viewpoints. So I totally agree that there is a need for education. There's a need for involvement from multiple stakeholders. But if I may, I think the companies we work with, they're still struggling with what does good look like. And this is where we believe that government has a critical role to play in thoughtful policymaking and in these standards to at least give that context to these companies because everyone right now, even if they're trying to self-regulate, do not know what does good look like. So our ask right now is really making sure that there is more transparency around how these systems are built and deployed.

Chairwoman STEVENS. Yes, right. And there's also certainly examples from throughout history where the notion of good has gotten it wrong.

But with that, why don't I turn it over to Mr. Feenstra, for 5 minutes of questioning. Thank you.

Mr. FEENSTRA. Thank you, Madam Chair. I'm so glad that we could have an extra round of questions.

And Dr. Isbell, thank you again for all your comments. I've been enjoying listening to you. And, as academics, to me, the challenge is—I finished my dissertation on maternity healthcare in rural America. And the challenge is, you know, we talk about ethics, but there's this fine line of how we access data and the barriers that are put on to try to get the data. And so how do we thread that needle of, you know, there's a need to have the data and to create trustworthy AI systems, and yet there's that balancing act of ethics. Can you dive into that a little bit?

Dr. ISBELL. I mean, I do have my opinions about how to solve all problems around ethics, which is a very deeply difficult question. I think the best way of thinking about it is to help people to articulate explicitly what it is that—what the tradeoffs are and where they want to live in that space of tradeoffs. If people can understand the tradeoffs, they can make informed decisions. I guarantee you that, first off, there's more data out about you out there in the world than you have ever imagined and that people know more about you than you wish that they did, and that could be a good thing because one day, it may save your life. On the other hand, it's a lot—it's your privacy, and it's who you are, and people shouldn't just be able to get access to that data just because they can.

Mr. FEENSTRA. Is there any data, though, that you'd say that would be beneficial that, you know, you look at and say, OK, this is captive that we can't get at that might be helpful as we move into trustworthiness and AI?

Dr. ISBELL. I think that that's a conversation that involves, as we've been saying all along, all the stakeholders who are involved.

I will add one thing, though, which is, although I think that bottom-up thinking is good and it's something that's driven us to innovation, it says right there in this chamber that, "Where there is no vision, the people perish."

Mr. FEENSTRA. That's right.

Dr. ISBELL. And the vision has to come from elected officials, it has to come from government, and it has to be a conversation about where it is we agree we want to go.

Mr. FEENSTRA. Yes, I agree. Thank you, very, very good and thoughtful words.

Ms. Singh, very intrigued by what your organization does. So if you look at how we build the appropriate safety and security into products, do you see a role in government? Or how do we incentivize going down this path, especially in the private sector? I mean, I think the private sector has some accountability in going down this path. But do you see anything that we can do? You know, we can put parameters, I get that. But we also, to me, have to do something to allow people to say I want to. Do you have any thoughts on that?

Ms. SINGH. Thank you so much for that question because I certainly do have many thoughts on it. But one that I would love to reemphasize here is the companies we work with right now, they are recognizing the importance of transparency reporting and disclosures because that transparency is helping them build trust with the consumers and truly get that competitive advantage. While one of the reasons that these companies are not sharing these transparency reports broadly is because they don't know how their competitors or others in the market stack up to it.

Mr. FEENSTRA. Yes.

Ms. SINGH. So at Credo AI, we are big proponents of you know, the government coming up with standards that cannot only mandate disclosures, but I think we will—it will propel a thoughtful benchmarking across these AI applications.

Mr. FEENSTRA. Yes, I mean, that's a great thought, that you can be protective in your data, but if we say—if the government says, wait a minute, this is universal data that everybody could use, that can be a gamechanger a little bit. Again, ethics plays a vital role in that. Thank you.

With that, I am out of time. Thank you.

Chairwoman STEVENS. Yes. And we'll hear from Dr. McNerney for 5 minutes of additional questioning.

Mr. MCNERNEY. Well, good. Now that you're back, I can thank you for having this hearing. It's great. And again, I want to thank the witnesses.

Ms. Singh, I feel bad about leaving you out first round, but I have two big concerns about AI, and I'll throw the first one to you. The first one is—and machine learning, which has really overtaken AI—that AI will overtake an increasing number of decisionmaking from humans, pushing us more and more into irrelevance and sort of dehumanizing us. What can we do to prevent that, you know, pushing us aside with the decisionmaking capability of AI?

Ms. SINGH. Thank you so much for that question. You know, with any disruptive technology, be it AI, we see there are huge economic impacts. And we see that in, you know, changes in work force, the

role that humans will play in the future of work. But as we step back and think about it, I think we have a great opportunity right now to invest more in education. As Dr. Isbell mentioned, I'm excited his son is going to be getting educated on AI because I think that's going to be critical. But thinking about reskilling and upskilling in this age of AI is going to give us a competitive edge.

Mr. MCNERNEY. So that's a great answer, educate more people so that we can utilize the AI in a more productive way than letting it make decisions for us. That's basically what you're saying, right?

Ms. SINGH. Yes, absolutely.

Mr. MCNERNEY. Very good. OK. Thank you.

The next one, I guess I'll go to Dr. Isbell again. AI—one of my other concerns about AI is that it's being used to monitor humans and our behaviors, our habits, especially either in autocratic nations or by businesses that would like to be able to influence our decisionmaking in terms of the way we spend our money. What do you think is a way to mitigate that issue?

Dr. ISBELL. So first off, you're right, that's exactly what happened, and it's been happening for a long time. Black Friday is a thing that happens because it gets people to buy things, right, so this is hardly new. What has happened is computing and AI has made it much more efficient and easier to deploy.

My answer to that—I have two. One is that it's education. It's making people aware of what's happening and allowing them to make reasonable decisions. The other is that there are policies and there are technical mechanisms that we can employ. We can encourage people to develop and to deploy that will allow them—that will allow people to understand what is being happened—what is happening to them. You are in fact being studied. You—your data is in fact predicting this behavior, and you're doing this. And giving people the tools, not just the stuff that they know—the education they learn on their own but the technical tools that allow others to monitor the monitors, that is a place that has a lot of potential and not one that we've invested a great deal into.

Mr. MCNERNEY. Well, the French postmodernists in the 1930's—1930's and 1940's were sort of warning us that the government would be getting more and more information about us and being able to use that information to control our political decisionmaking as individuals, and that's sort of what I was worried about. And now what we're seeing with social media is that these—some of these companies are using information to direct people into political bubbles that may advocate violence or other sorts of extreme behavior. And I think that's one of the issues we have—that I'm having with how do we tamp that down? Do you have any recommendations, Mr. Crenshaw, on how we could go about doing that?

Mr. CRENSHAW. Well, I think when it comes to anytime we're looking at the use of algorithms, we have to look at it from a risk-based approach. And I think we also need to realize that there are some benefits also to artificial intelligence that we've seen. And, you know, one of the things I wanted to note is that what we've learned is that the more people know about AI, the less scared or concerned they are about it. And I think that's why education about artificial intelligence is so important. But companies also

need to build in ethics and ethical decisionmaking into their AI as well, too. And we see companies that are leading in this space.

Mr. MCNERNEY. But it's hard to regulate that. And I'm thrilled that we're hearing about companies hiring ethicists, but how do we get that as a part of the corporate mindset that, you know, we need to do this in the future? So—it's not something we can regulate I don't think.

Mr. CRENSHAW. I agree that C-suite needs to be involved. It needs to be part of corporate culture is building in ethics into artificial intelligence. But at the same time, I think with the work we're seeing at agencies like NIST are getting us in the right direction toward where we want to be.

Mr. MCNERNEY. Thank you. I yield back.

Chairwoman STEVENS. Thank you. And with that, I don't believe we have any other questions. So we're going to bring the hearing to a close. Do we have one more? Oh, did Baird come back? OK, hold on. I'm not closing. Where is he? Dr. Baird? He's not coming? Well, we got questions for the record, too. OK. We're prepared to close. All right. Well, we're prepared to close. But honestly, we're not going to close the door on the conversation because this has only brought up more questions. And in fact, we could probably have a hearing on a couple of different subsets that we discussed today. I believe with this Committee, and as Mr. Gonzalez who, you know, we have been so privileged to work with during his couple of terms here in the Congress, mentioned, you know, taking research applications, commercializing them, recognizing where our economy filters in.

We also recognize that we're in a leadership moment, and this is—you know, we have been deeply privileged to have Dr. McNerney through his tenure, his mighty tenure in the Congress on this Committee, and he's so, so dedicated to this Committee, but this is a leadership moment for the United States of America. And we are going to shape how the world's going to go on this. We want to be able to shape how the world's going to go, and we've got to be prepared to do some of the deeper work. It's not just the question of harm, but it's also the questions of, you know, the meta challenges that come before us that are somewhat brought on by AI. It's forcing us to be more collaborative. It is forcing us to come together in ways that we didn't last century.

I left out that I was working at a digital research lab before coming to this body and we did the taxonomy, Mr. Crenshaw, on the IoT (Internet of things) jobs, you know, how companies are going to have to hire. We did this in partnership with Manpower Group and a host of other industry and academic partners. Digital ethicists came up. That was one of the job profiles we came up with. That was just 5, 6 years ago. And I mentioned Turing test, and we were so possessed when I was in school by the Turing test, like we thought that was going to be the question. And Mr. Sherman sort of got to that in his questions, you know, are we worried about replacing humanity? No, we are talking about what Mr.—or Dr. Isbell said in his testimony, culture, changing culture and how we influence culture through the laws we pass in this body.

And we have been addressing some meta challenges. I didn't have the privilege of having Mr. Feenstra here last term, but I

know we would have been working together on the trade deal, the USMCA (United States-Mexico-Canada Agreement). You have unions and the Chamber came together to pass USMCA. This time around, we passed *Inflation Reduction Act*. For the first time ever, you know, we're dealing with climate. You've got the environmental groups and the industry partners, my automakers saying they want the same thing.

So these digital applications, these complex artificial intelligence systems that we're putting into place, they're asking us to come together. So, Ms. Tabassi, I—you know, we're going to come back to you because we just—we think NIST solves all of our problems, the mighty agency that can with a little. And we're excited about that, and we—and we're going to come visit you and we're going to talk about how you're stitching together with your risk management what Dr. Isbell said and what Ms. Singh is saying. Who's at the table? Who's at the table? You know, we solve some problems in ones and twos, and then we look at some of the broader challenges. But overall, we're wildly optimistic. We're working on the vision, and we're excited that we had this time together today. Hopefully, the rest of the Congress tunes in on C-SPAN later.

But with that, we're going to close it. We're going to leave the record open for a couple of weeks for additional questions for the record, and our witnesses are excused. Thank you.

[Whereupon, at 12:29 p.m., the Subcommittee was adjourned.]

Appendix I

ANSWERS TO POST-HEARING QUESTIONS

ANSWERS TO POST-HEARING QUESTIONS

Responses by Ms. Elham Tabassi

U.S. HOUSE OF REPRESENTATIVES
COMMITTEE ON SCIENCE, SPACE, AND TECHNOLOGY
SUBCOMMITTEE ON RESEARCH AND TECHNOLOGY

“Trustworthy AI: Managing the Risks of Artificial Intelligence”

Questions for the Record to:

Ms. Elham Tabassi
Chief of Staff, Information Technology Laboratory
National Institute of Standards and Technology
Submitted by Representative Dan Kildee

1. **Representative Dan Kildee:** I understand that NIST has sought comments on a second draft of the Artificial Intelligence Risk Management Framework. It is critical that, as AI technologies develop, the inputs used to train AI systems are procured in a way that respects copyright and other intellectual property rights. Will NIST act on public comments calling for better safeguards around the use of protected works as training data for AI systems, including identification of IP rights implicated by training materials and all necessary information to ensure creators are compensated for use of their works?

NIST Response:

NIST is integrating public comments into the AI Risk Management Framework that reflect a respect for copyright and other intellectual property rights. This includes concerns around training data that may have been sourced from copyrighted works, or that is otherwise subject to third-party intellectual property rights. NIST is on track to publish version 1.0 of the AI Risk Management Framework in early 2023 consistent with congressional direction. Throughout the process of developing the Framework, NIST has engaged extensively with relevant stakeholders through public comment, workshops, and discussions.

Responses by Mr. Jordan Crenshaw



U.S. Chamber of Commerce

1615 H Street, NW
Washington, DC 20062-2000
uschamber.com

November 3, 2022

The Honorable Haley Stevens
Chair
Subcommittee on Research
and Technology
Committee on Science,
Space and Technology
U.S. House of Representatives
Washington, DC 20515

The Honorable Randy Feenstra
Ranking Member
Subcommittee on Research
and Technology
Committee on Science,
Space and Technology
U.S. House of Representatives
Washington, DC 20515

Dear Chairwoman Stevens and Ranking Member Feenstra:

Please see below my response to the Question for the Record I received after my testimony in front of your Subcommittee for the September 29 hearing, "Trustworthy AI: Managing the Risks of Artificial Intelligence."

Question: In October 2022, the White House issued the *Blueprint for an AI Bill of Rights*, which is comprised of "a set of five principles and associated practices to help guide the design, use, and deployment of automated systems to protect the rights of the American public in the age of artificial intelligence." From the U.S. Chamber's perspective, how does the Administration's approach through the AI Bill of Rights differ from NIST's approach through the AI Risk Management Framework to bolster innovation and transparency in trustworthy AI? Do you find the Administration's approach helpful to building consumer confidence and trustworthiness of AI systems?

OSTP's and administrations' release of the AI Bill of Rights does not help foster innovation, transparency, and trustworthiness in AI. Furthermore, we see that the release of the AI Bill of Rights creates potentially significant conflicts with the congressionally mandated NIST AI Risk Management Framework for the following reasons.

I. Creating Uncertainty and Conflicting Frameworks

The "blueprint" puts unnecessary uncertainty in the current domestic and international work taking place to develop "trustworthy AI." As the United States continues to develop and refine the Congressionally mandated NIST Risk

Management Framework, and our international counterparts such as Israel, Singapore, Japan, Canada, and the EU continue to look to work on these matters, it is essential for the United States government to lead by example with one specific plan to help move domestic and international policy in a direction which helps American businesses, continue innovation, and to allow for the opportunity to address trustworthy AI holistically and thoughtfully.

The AI Bill of Rights has done the exact opposite, as our allies are now confused about the stance and direction the United States is taking on current and future policy around the development of AI. For example, it has been reported¹ that the United States recently provided critical feedback to the EU on its forthcoming EU AI Act. That feedback includes the United States support for “individualized risk assessments.”² However, this explicitly contradicts the AI Bill of Rights, which advocates for a broader regulatory approach.

II. Lack of Transparency

The adoption of the AI Bill of Rights was not a transparent process, which harms businesses and the American public’s trust to be a part of these critical conversations. Although the “Blueprint” highlights Organizations from which OSTP met and received feedback, we would like to emphasize that the process lacked the openness and transparency necessary to obtain sufficient stakeholder input about these complex issues. Furthermore, the only request for information from OSTP regarding the “AI Bill of Rights” was related to biometrics³ and not artificial intelligence. Without the necessary stakeholder feedback on matters the blueprint addresses, OSTP fails to create a complete record of the use of the technology. This contradicts the work that has transpired at NIST with the Risk Management Framework. NIST, at this time, has done three workshops and four RFIs around the development of the RMF. The timing of the release of the Bill of Rights, given its transparency deficiencies, is disappointing in light of the congressionally mandated stakeholder driven approaches at NIST and the National AI Advisory Committee.

III. Unworkable Definitions

¹ <https://www.euractiv.com/section/digital/news/the-us-unofficial-position-on-upcoming-eu-artificial-intelligence-rules/>

² <https://www.euractiv.com/section/digital/news/the-us-unofficial-position-on-upcoming-eu-artificial-intelligence-rules/>

³ <https://www.federalregister.gov/documents/2021/10/08/2021-21975/notice-of-request-for-information-rfi-on-public-and-private-sector-uses-of-biometric-technologies>

Definitions within the blueprint do not help harmonization. While defining terms is a critical step, the definitions used within the “Blueprint” could potentially harm the United States’ ability to identify the appropriate and necessary lexicon among like-minded international allies. For example, the definition of “Automated System” is comprehensive and the use of the phrase “includes, but not limited to,” leads to unnecessary uncertainty around what is an “Automated System.” Any definition of an Automated System must be clearly defined. This counter to the NIST Framework uses a Congressionally enacted definition, which aligns with the OECD’s AI Principle.

IV. Concerns with Audits

The Blueprint’s call for independent evaluations by third-party auditors also raises concerns. There are no concrete standards and metrics for auditing Artificial Intelligence systems. The Blueprint’s call to allow “Independent Evaluators, such as...journalists...third-party auditors” to be “given...unfiltered access to the full system” is pointless at a time when independent evaluations of AI systems continue to lack any standardization. NIST Risk Management Framework runs counter to this part of the blueprint, as the Framework is about developing internal consciences about addressing and mitigating bias instead of opening businesses and organizations to unfiltered access. Furthermore, NIST is producing the upcoming playbook, which provides suggested actions, references, and documentation guidelines for stakeholders to achieve outcomes⁴.

V. Conflating Data Privacy with Algorithmic Policy

The blueprint conflates data-privacy with Artificial Intelligence: The Blueprint lists “Data Privacy” as one of the five principles of the Blueprint. While we wholeheartedly agree that data is a significant part of Artificial Intelligence, it is essential to highlight that the two are distinctly different issues. Data Privacy has long been understood to be how an individual’s data is used and shared. Where Artificial Intelligence is when the data is used in conjunction with algorithms that learn from that data to do a specific assigned task, it is essential not to conflate these two issues, as the nuances and complexities in each case are distinctly different.

The Chamber also takes exception with the term “surveillance” when referring to the use of data broadly, as the A.I. Bill of Rights appears to do. In its current Advanced Noticed of Proposed Rulemaking related to “commercial surveillance,” the FTC utilizes a definition of commercial surveillance that effectively captures all data analysis in business.⁵ This term is used pejoratively without considering technology’s benefits for

⁴ <https://pages.nist.gov/AIRMF/>

⁵ 87 Fed. Reg. 51277 “For the purposes of this ANPR, “commercial surveillance” refers to the collection, aggregation, analysis, retention, transfer, or monetization of consumer data and the direct derivatives of that

things like the affordability of goods and services, financial inclusion, public safety, and improving health outcomes.

VI. Call For Codification

The Chamber is deeply concerned that the “blueprint” intends to influence state and local government to model legislation after its principles and recommendations. This was stated as one of the goals in a blog post by OSTP during the reveal of the blueprint, which stated that “policymakers can codify these measures into law or use the framework and its technical companion to help develop specific guidance on the use of automated systems within a sector.”⁶

The call to codify principles that have not been fully vetted, discussed and analyzed on their specific merits and economic impact will lead to unintended consequences for those communities. The use of the blueprint to validate efforts to regulate the use of Algorithms is already occurring. For example, the Attorney General of the District of Columbia authored a blog post highlighting that he “supports the white house AI Bill of Rights⁷,” which “includes Core Aspects of His Office’s Bill⁸.” It is essential that communities have the necessary conversation and dialogue about using technology to build understanding and trust. Instead, the codification of the AI Bill of Rights could lead to unnecessary regulations, which never received the necessary discussion and analysis.

Sincerely,



Jordan Crenshaw
Vice President
Chamber Technology Engagement Center
U.S. Chamber of Commerce

information. These data include both information that consumers actively provide—say, when they affirmatively register for a service or make a purchase—as well as personal identifiers and other information that companies collect, for example, when a consumer casually browses the web or opens an app. This latter category is far broader than the first.”

⁶ <https://www.whitehouse.gov/ostp/news-updates/2022/10/04/blueprint-for-an-ai-bill-of-rights-a-vision-for-protecting-our-civil-rights-in-the-algorithmic-age/>

⁷ <https://oag.dc.gov/release/ag-racine-supports-white-house-ai-bill-rights>

⁸ <https://oag.dc.gov/release/ag-racine-supports-white-house-ai-bill-rights>

Appendix II

ADDITIONAL MATERIAL FOR THE RECORD

DOCUMENT SUBMITTED BY REPRESENTATIVE BRAD SHERMAN

Engineered Intelligence: Creating a Successor Species

Congressman Brad Sherman
House Floor Speech – May 17, 2000
(Updated on May 17, 2019)

I believe that the impact of science on this century will be far greater than the enormous impact science had on the last century.

As futurist Christine Peterson notes: "If someone is describing the future 30 years from now and they paint a picture that seems like it is from a science fiction movie, they might be wrong. But, if someone is describing the future a generation from now and they paint a picture that doesn't look like a science fiction movie, then you know they are wrong." We are going to live in a science fiction movie, we just don't know which one.

There is one issue that I think is more explosive than even the spread of nuclear weapons: engineered intelligence. By that I mean, the efforts of computer engineers and bio-engineers who may create intelligence beyond that of a human being.

In testimony at the House Science Committee¹, the consensus of experts testifying was that in roughly 25 years we would have a computer that passed the Turing Test², and more importantly, exceeded human intelligence.

As we develop more intelligent computers, we will find them useful tools in creating ever more intelligent computers, a positive feedback loop. I don't know whether we will create the maniacal Hal from *2001*, or the earnest Data from *Star Trek* --- or perhaps both.

There are those who say don't worry, even if a computer is intelligent and malevolent --- it is in a box and it cannot affect the world. But I believe that there are those of our species who sell hands to the Beelzebub, in return for a good stock tip.

I do draw solace from the fact that just because a computer is intelligent, or even self-aware, this does not mean that it is ambitious.

By ambitious, I mean possessing a survival instinct together with a desire to affect the environment so as to ensure survival, and often a desire to propagate or expand.

My washing machine does not seem to care whether I turn it off or not. My pet mouse does seem to care. So even a computer possessing great intelligence may simply

have no ambition, survival instinct, or interest in affecting the world.

"We are going to live in a science fiction movie, we just don't know which one."

DARPA³ is the government agency on the cutting edge of supercomputer research. I have urged DARPA to develop computer systems designed to maximize the computer's utility, while avoiding self-awareness, or at least ambition.

Bio-engineers may be able to start with human DNA and create a 2,000 pound mammal with a 300 pound brain designed to beat your grandkids on the LSAT. No less troubling, they might start with canine DNA and create a mammal with sub-human intelligence, and no civil rights.

DNA is inherently ambitious. Those microbes which didn't seek to survive or replicate, didn't. Birds seem to care whether they or their progeny survive, and they seek to affect their environment to achieve that survival.

In any case, you have the bio-engineers and the computer engineers both working toward new levels of intelligence. I believe in our lifetime we will see new species

possessing intelligence which surpasses our own.

The last time a new higher level of intelligence arose on this planet was roughly 50,000 years ago. It was our own ancestors, who then said hello to the previously most intelligent species, Neanderthals. It did not work out so well for the Neanderthals.

I used to view this as a contest between the bio-engineers and the computer engineers (or if you use the cool new lingo, wet nanotechnology and dry nanotechnology), in an effort to develop a new species of superior intelligence. I felt that the last decision that humans would make would be whether our successors are carbon-based or silicon-based;⁴ the product of bio-engineering or of computer engineering.

Now I believe we are most likely to see combinations that will involve nature, computer engineering, and bio-engineering: humans with pharmaceutical intelligence boosters; DNA enhancements; computer-chip implants; or all three.

First, this will be used to cure disease, then to enhance human capacity. The enhanced-human will precede the trans-human.

“Will our successors be carbon-based or silicon-based?”

Now how should we react to all of this? It is important that we benefit from science, even as we consider its more troubling implications.

I chaired the House Subcommittee on Nonproliferation which deals with the only other technologies that pose an existential threat to humankind, namely the proliferation of nuclear and biological weapons.

The history of nuclear technology is instructive. On August 2, 1939, Einstein sent Roosevelt a letter saying a nuclear weapon was possible; six years later, nuclear technology literally exploded onto the world scene. Only after society saw the negative effects of nuclear technology, did we see the prospects for nuclear power and nuclear medicine.

The future of engineered intelligence will be different. The undeniable benefits of computer and DNA research will arrive long before the problematic possibilities. Their introduction will be gradual, not explosive. Fortunately, we will have far more than six years to consider the implications --- unless we choose to squander the next few decades. My fear is that our philosophers, ethicists and society at large, will ignore the issues that will

inevitably present themselves until they actually present themselves. And these issues require more than a few years of thought.⁵

I am confident that if we plan ahead we can obtain the utility of supercomputers, and the benefits of bio-engineering, without creating new levels of intelligence. We can then pause and decide whether we in fact wish to create a new intelligent species or two.

Finally, I would quote Oliver Wendell Holmes who said 100 years ago, “I think it not improbable that man, like the grub that prepares a chamber for the winged thing it never has seen but is to be -- that man may have cosmic destinies that he does not understand.”⁶

Likewise, it is possible that our grandchildren --- or should I say “our successors” --- will have less resemblance to us than a butterfly has to a caterpillar. Our best minds in philosophy, science, ethics and theology ought to be focused on this issue.

1. On April 9, 2003, the U.S. House Committee on Science and Technology, held a hearing titled “The Societal Implications of Nanotechnology.”

2. If a human receives a text message and cannot determine if it was composed by a computer or a human, then the computer has passed the Turing Test.

3. The Defense Advanced Research Projects Agency (DARPA).

4. Despite the fact that supercomputers may not use chips with silicon substrate, for these purposes, we'll still refer to computer chips as “silicon.”

5. This issue is discussed in “Drive New World War” by Jamie Metz, Published in Issue 8, Spring 2008, Democracy: A Journal of Ideas.

6. Oliver Wendell Holmes, “Law and the Court,” speech at the Harvard Law School Association of New York, 15 February 1913.